

Published in final edited form as:

J Theor Biol. 2012 February 7; 294: 153–171. doi:10.1016/j.jtbi.2011.10.032.

Confinement of gene drive systems to local populations: A comparative analysis

John M. Marshall^{a,b} and Bruce A. Hay^a

^aDivision of Biology, California Institute of Technology, Pasadena, California 91125, USA

^bMRC Center for Outbreak Analysis & Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, W2 1PG, UK

Abstract

Mosquito-borne diseases such as malaria and dengue fever pose a major health problem through much of the world. One approach to disease prevention involves the use of selfish genetic elements to drive disease-refractory genes into wild mosquito populations. Recently engineered synthetic drive systems have provided encouragement for this strategy; but at the same time have been greeted with caution over the concern that transgenes may spread into countries and communities without their consent. Consequently, there is also interest in gene drive systems that, while strong enough to bring about local population replacement, are unable to establish themselves beyond a partially-isolated release site, at least during the testing phase. Here, we develop simple deterministic and stochastic models to compare the confinement properties of a variety of gene drive systems. Our results highlight several systems with desirable features for confinement – a high migration rate required to become established in neighboring populations, and low-frequency persistence in neighboring populations for moderate migration rates. Single-allele underdominance and single-locus engineered underdominance have the strongest confinement properties, but are difficult to engineer and require a high introduction frequency, respectively. Toxin-antidote systems such as *Semele*, *Merea* and two-locus engineered underdominance show promising confinement properties and require lower introduction frequencies. Killer-rescue is self-limiting in time, but is able to disperse to significant levels in neighboring populations. We discuss the significance of these results in the context of a phased release of transgenic mosquitoes, and the need for characterization of local ecology prior to a release.

Keywords

dengue fever; difference equation; malaria; migration; population replacement; stochastic process; transgenic mosquitoes; underdominance

© 2011 Elsevier Ltd. All rights reserved.

Corresponding author: John M. Marshall, Department of Infectious Disease Epidemiology, Division of Epidemiology, Public Health and Primary Care, Imperial College London, London, W2 1PG, UK, Voice: +44-207-594-1379, Fax: +44-207-262-3180, john.marshall@imperial.ac.uk.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Mosquito-borne diseases such as dengue fever, chikungunya and malaria continue to pose a major health problem through much of the world. Treatments for dengue fever and chikungunya remain elusive, and malaria is proving exceptionally difficult to control in highly-endemic areas with insecticide-treated nets, indoor residual spraying and antimalarial drugs (Griffin *et al.*, 2010; WHO, 2010). The failure of currently available methods to control these diseases has renewed interest in approaches to disease prevention that involve the use of gene drive systems to spread disease-refractory genes into wild mosquito populations (Alphey *et al.*, 2002; Marshall & Taylor, 2009). Such strategies are attractive because they are self-perpetuating and are expected to result in disease suppression far beyond the release site.

A number of gene drive systems have been proposed, including naturally-occurring selfish genetic elements such as transposable elements (TEs), post-meiotic segregation distorters, *Medea* elements, homing endonuclease genes (HEGs), and the intracellular bacterium *Wolbachia* (Braig & Yan, 2001; Gould & Schliekelman, 2004; Sinkins & Gould, 2006). Another set of approaches to bringing about population replacement involves creating insects in which genes of interest are linked to engineered chromosomes: compound chromosomes or translocations (Curtis, 1968; Foster *et al.*, 1972), or pairs of unlinked lethal genes, each of which is associated with a repressor of the lethality induced by expression of the other lethal gene – a system known as engineered underdominance (Davis *et al.*, 2001; Magori & Gould, 2006).

An essential feature of population replacement is the ability of released transgenic insects to spread disease-refractory genes through a wild population on a human timescale. The observation that several naturally-occurring selfish elements have spread over wide geographical areas – a TE in *Drosophila melanogaster* (Kidwell, 1983), *Wolbachia* in *Drosophila simulans* (Turelli & Hoffmann, 1991), and *Medea* in *Tribolium* (Beeman & Friesen, 1999; Lorenzen *et al.*, 2008) – provides encouragement for this strategy. However, these observations also highlight the potential for gene drive systems to spread transgenes into countries before they have agreed to their introduction (Knols *et al.*, 2007). The Cartagena Protocol – the United Nations protocol on the international movement of genetically modified organisms (GMOs) – is prohibitive of a release of GMOs capable of self-propagating across national borders in the absence of a multilateral international agreement (Marshall, 2010). It also allows nations to decide for themselves how they would like to regulate the import of GMOs.

The movement of mosquitoes carrying transgenes also requires consideration at the community level. A recent survey of public attitudes to a release of malaria-refractory mosquitoes in Mali suggests that a small number of people would be happy for the first release to be conducted in their community; however, most people would first like to see the results of a release in an isolated community where malaria prevalence has been shown to decrease in the absence of side-effects (Marshall *et al.*, 2010). These regulations and societal views highlight the importance of confining the first releases of transgenic mosquitoes to isolated locations, particularly when gene drive systems are involved. This is in agreement with the principle of scientific risk management, which recommends that the magnitude of potential risks be minimized by spatially limiting the release and providing mechanisms for removal of transgenes in the event of adverse effects (Beech *et al.*, 2009).

Consequently, the strategy of population replacement is left with two competing mandates, at least during the testing phase. Transgenes must be able to spread to high frequencies locally, in order to provide an adequate test of the technology's ability to prevent disease

(Boete & Koella, 2002; Boete & Koella, 2003); but they must not spread to significant levels in neighboring populations. The use of trap crops, vegetation-free zones and spatial isolation have been discussed in the context of confining an accidental release of transgenic mosquitoes from field cages (Benedict *et al.*, 2008). Additionally, studies of mosquito ecology on islands off the coast of Africa have begun with an isolated transgenic release in mind (Pinto *et al.*, 2003). These safeguards are effective at reducing the migration rate of mosquitoes to nearby locations; however, a small number of escapees are inevitable and, if these escapees carry invasive gene drive systems, this may be sufficient for transgenes to spread beyond their release site. The release of mosquitoes carrying only transgenes conferring disease refractoriness has been proposed (Benedict *et al.*, 2008) and is an important step towards assessing the behavior of the refractory gene in the wild. However, gene drive systems will ultimately be required to achieve the transgenic frequencies necessary for disease control without prohibitive release sizes.

Underdominant systems provide a potential solution to these competing mandates (Altrock *et al.*, 2010). The mosquito species that transmit human diseases are largely anthropophilic, forming subpopulations around human settlements as discrete blood sources (Service, 1993). Abstract two-population models have shown that an underdominant allele can become established in one population while persisting in a neighboring population at low frequencies provided that the migration rate between populations is sufficiently low and the strength of selection against heterozygotes is sufficiently strong (Karlin & McGregor, 1972; Lande, 1985; Altrock *et al.*, 2010). The single underdominance allele described in these models is particularly difficult to engineer; however several genetic systems displaying similar properties have been proposed, including translocations (Curtis, 1968), engineered underdominance (Davis *et al.*, 2001), and a variety of single-locus toxin-antidote systems such as *Semele* (Marshall *et al.*, 2011), inverse *Medea* (Marshall & Hay, 2011a) and *Medea* with a recessive antidote (*Merea*).

A proper assessment of the ability to confine these systems to discrete populations will require a detailed ecological analysis, taking into account phenomenological features of the mosquito populations of interest – including seasonally-fluctuating population sizes and migration rates as well as seasonally-varying chromosomal form and species make-ups (Lanzaro *et al.*, 1998; Taylor *et al.*, 2001; Tripet *et al.*, 2005). The present analysis, however, focuses on a comparison of the gene drive systems currently being considered, and their relative ability to be confined to partially-isolated populations. For the purposes of comparison, we consider the simplest possible model of population structure – a two-population model in which mosquitoes with gene drive systems are introduced into one population, which exchanges migrants with a neighboring population. This differs from the analysis of Marshall (2009), which uses a branching process framework to compare the ability of gene drive systems to persist in the environment following an accidental release from a contained facility. Here, we use a simple difference equation framework to see whether there is a realistic set of parameters for which a transgene can become established at its release site without spreading into neighboring populations. We discuss the relevance of these findings to a phased release of transgenic mosquitoes intended for the control of vector-borne diseases, and to the strategy of population replacement in general.

2. Model development

We use two modeling frameworks to compare the degree to which mosquitoes engineered with various gene drive systems can be confined to their release site. First, we consider a source model in which transgenic mosquitoes have already reached equilibrium in population A, and population A donates a fraction, μ , of its population to population B at each generation (Fig. 1A). We include this model for its analytic tractability, and because it

provides a good first approximation of a model in which migration occurs in both directions. Second, we consider a metapopulation model consisting of two populations, each of which exchanges a fraction, μ , of its population with the other at each generation (Fig. 1B). The transgenic μ mosquitoes are introduced into population A, while population B initially consists of wild-types. In both cases, we assume discrete generations and random mating. The inheritance pattern of each gene drive system is then used to calculate genotype frequencies in the next generation for both populations.

We illustrate these two models, using *Medea* as a case study, in the following section. We then adapt these models to incorporate the inheritance patterns of invasive gene drive systems (§4) such as HEGs (Deredec *et al.*, 2008), TEs (Charlesworth *et al.*, 1994) and *Wolbachia* (Turelli & Hoffmann, 1999); underdominant systems (§5) such as translocations (Curtis, 1968), engineered underdominance (Davis *et al.*, 2001) and an underdominant allele (Altrock *et al.*, 2010); other non-invasive systems (§6) such as *Semele* (Marshall *et al.*, 2011), inverse *Medea* (Marshall & Hay, 2011a) and *Merea* (Marshall, 2011); and killer-rescue (§7; Gould *et al.*, 2008). Finally, we adapt the two-population model to account for a series of five populations to monitor the spread of a gene drive system beyond an immediately neighboring population.

3. Medea

Medea elements spread through natural populations by causing the death of all offspring of heterozygous females that do not inherit the allele (Beeman *et al.*, 1992; Wade & Beeman, 1994). Synthetic *Medea* elements have been engineered in *Drosophila* by linking a maternally-expressed toxin with a zygotic antidote (Fig. 1C), and have been shown to drive population replacement in laboratory cage populations (Chen *et al.*, 2007). We consider a *Medea* element as a single allele, which we denote by “*M*,” and refer to the absence of the *Medea* allele at the corresponding locus as “*m*.” In population A, the proportion of the *k*th generation that are mosquitoes of genotypes *mm*, *Mm* and *MM* are denoted by $u_{A,k}$, $v_{A,k}$ and $w_{A,k}$, respectively. The corresponding proportions for population B are $u_{B,k}$, $v_{B,k}$ and $w_{B,k}$.

We begin by considering the population dynamics of the *Medea* allele without migration, assuming random mating and infinite population size (Ward *et al.*, 2010). By considering all possible mating pairs, and by taking into account that wild-type offspring of heterozygous mothers are unviable, the genotypes of embryos in the next generation in population A are described by the ratio $\hat{u}_{A,k+1} : \hat{v}_{A,k+1} : \hat{w}_{A,k+1}$, where,

$$\hat{u}_{A,k+1} = u_{A,k}^2 + 0.5u_{A,k}v_{A,k}, \quad (1)$$

$$\hat{v}_{A,k+1} = 2u_{A,k}w_{A,k} + 0.5v_{A,k}^2 + u_{A,k}v_{A,k} + w_{A,k}v_{A,k}, \quad (2)$$

$$\hat{w}_{A,k+1} = w_{A,k}^2 + w_{A,k}v_{A,k} + 0.25v_{A,k}^2. \quad (3)$$

Here, we have assumed 100% toxin efficiency, which has been achieved by Chen *et al.* (2007). Normalizing these ratios and taking into account fitness costs, the genotype frequencies in the next generation are given by:

$$u_{A,k+1} = \widehat{u}_{A,k+1} / W_{A,k+1}, \quad (4)$$

$$v_{A,k+1} = \widehat{v}_{A,k+1} (1 - hs) / W_{A,k+1}, \quad (5)$$

$$w_{A,k+1} = \widehat{w}_{A,k+1} (1 - s) / W_{A,k+1}. \quad (6)$$

Here, s and hs represent the fitness costs associated with being homozygous or heterozygous for the element, and W is a normalizing term given by,

$$W_{A,k+1} = \widehat{u}_{A,k+1} + \widehat{v}_{A,k+1} (1 - hs) + \widehat{w}_{A,k+1} (1 - s). \quad (7)$$

Analogous equations apply to population B.

Source model

For the source model, we begin by assuming that the transgene has already reached equilibrium in population A, ignoring migratory effects. We calculate this equilibrium by solving the equality,

$$(u_{A,k+1}, w_{A,k+1}) = (u_{A,k}, w_{A,k}) = (u_{A,*}, w_{A,*}), \quad (8)$$

and choosing the stable equilibrium with the highest transgenic allele frequency. To determine whether the equilibrium is stable or not, we calculate the eigenvalues of the Jacobian matrix,

$$\left(\begin{array}{cc} \frac{\partial u_{k+1}}{\partial u_k} & \frac{\partial u_{k+1}}{\partial w_k} \\ \frac{\partial w_{k+1}}{\partial u_k} & \frac{\partial w_{k+1}}{\partial w_k} \end{array} \right) \bigg|_{(u_k, w_k) = (u_*, w_*)}. \quad (9)$$

The equilibrium is locally stable if all eigenvalues have modulus less than one, and is unstable if one or more of the eigenvalues have modulus greater than one (Elaydi, 1995). Otherwise, the analysis is inconclusive. For simplicity, let us assume a fitness cost with a heterozygosity of $h = 0.5$. This allows us to describe the dynamics of the *Medea* allele by a single parameter – the homozygous fitness cost, s . Solving the equality in eqn (8) then gives us four solutions,

$$(u_{A,*}, w_{A,*}) = \left\{ (0, 1), (1, 0), (0, 1 - s), \left(\frac{4 + s(s - 6)}{4 - 2s}, \frac{s^2}{4 - 2s} \right) \right\}. \quad (10)$$

The first of these represents fixation, the second represents loss, the third represents a mixture of heterozygotes and homozygotes, and the fourth represents a mixture of all three

genotypes. Calculating the stabilities of these equilibria, we see that loss is stable, fixation is stable in the absence of a fitness cost and unstable in the presence of a fitness cost, and the third equilibrium is stable for fitness costs less than $3 - \sqrt{5}$ and unstable for fitness costs greater than $3 - \sqrt{5}$. The fourth equilibrium is only biologically relevant for fitness costs less than $3 - \sqrt{5}$ and is unstable over this entire range. This latter solution represents one of a family of threshold points for fitness costs less than $3 - \sqrt{5}$ (0.764), above which the *Medea* allele spreads to an equilibrium frequency of $(u_{A,*}, w_{A,*}) = (0, 1 - s)$, and below which it is lost.

We assume a fitness cost of less than 0.764 and a super-threshold release in population A, leading to the following genotype equilibrium in the source population,

$$(u_A, w_A) = (0, 1 - s). \quad (11)$$

Ignoring migratory effects in the source population, this equilibrium is maintained and population A donates mosquitoes with this set of genotype frequencies to population B at a rate, μ , measured relative to the size of population B. This allows us to calculate the genotype frequencies in population B analogously to those in population A with a constant influx of homozygotes at a rate of $w_A\mu = (1 - s)\mu$ per generation, and a constant influx of heterozygotes at a rate of $v_A\mu = s\mu$ per generation. Eqns (1–3) remain unchanged for population B, and eqns (4–6) become,

$$u_{B,k+1} = \widehat{u}_{B,k+1} / (W_{B,k+1} + \mu), \quad (12)$$

$$v_{B,k+1} = \widehat{v}_{B,k+1} (1 - hs) + s\mu / (W_{B,k+1} + \mu), \quad (13)$$

$$w_{B,k+1} = \widehat{w}_{B,k+1} (1 - s) + (1 - s)\mu / (W_{B,k+1} + \mu). \quad (14)$$

The normalizing term, $W_{B,k+1}$, is equivalent to that in eqn (7).

We are interested in the frequency that the introduced allele reaches in population B, as a function of its fitness cost and immigration rate. Most of all, we are interested in the conditions under which the allele can spread in population A without spreading into population B. To characterize the dynamics of the *Medea* allele in population B, we begin by calculating its equilibria by solving the equality,

$$(u_{B,k+1}, w_{B,k+1}) = (u_{B,k}, w_{B,k}) = (u_{B,*}, w_{B,*}). \quad (15)$$

This gives us five solutions,

$$(u_{B,*}, w_{B,*}) = \left\{ \left(0, \pm \sqrt{1+4\mu}, (0, 1-s), \left(\frac{\pm 4z(1+\mu) + (8 \mp 5z + 2\mu(4 \mp z))s - (24 \mp z + 24\mu)s^2 + (25+26\mu)s^3 - 2(5+6\mu)s^4 + (1+2\mu)s^5}{\frac{2s(2-s(3-s))(2(1+\mu)-s(2+\mu))}{s(1-s)(s-2\mu(2-s)) \mp z}}, \right) \right\}, \quad (16)$$

where $z = \sqrt{s^2(1-s)^2(s^2 - 4\mu(2-s)^2)}$. The first two of these (the upper and lower signs of the first entry) are not biologically relevant, the third represents a mixture of homozygotes and heterozygotes, and the fourth and fifth represent a mixture of all genotypes. Calculating the stabilities of these equilibria, we see that equilibrium three is stable, and the fourth and fifth equilibria only exist when $\mu \leq s^2/(4-2s)^2$. If this condition is satisfied, then the fourth equilibrium (the upper sign of the third entry) is stable and the fifth equilibrium (the lower sign) is unstable. The location and stability of the three biologically relevant equilibria suggests that, if the immigration rate is greater than $s^2/(4-2s)^2$, then the *Medea* element will reach the same prevalence in population B as in the source population; but if the immigration rate is smaller than $s^2/(4-2s)^2$, then the *Medea* element will persist in population B at a low level. In summary,

$$(u_{B,*}, w_{B,*}) = \begin{cases} \left(\frac{4z(1+\mu) + (8 - 5z + 2\mu(4 - z))s - (24 - z + 24\mu)s^2 + (25+26\mu)s^3 - 2(5+6\mu)s^4 + (1+2\mu)s^5}{\frac{2s(2-s(3-s))(2(1+\mu)-s(2+\mu))}{s(1-s)(s-2\mu(2-s)) - z}}, \right), & \mu \leq \frac{s^2}{(4-2s)^2} \\ (0, 1-s), & \mu > \frac{s^2}{(4-2s)^2} \end{cases} \quad (17)$$

The expression, $s^2/(4-2s)^2$, therefore represents a migration threshold for *Medea* under the source model – above this migration rate, the *Medea* allele spreads into population B; and below this migration rate, the *Medea* allele persists at a low level in population B. To put this expression into perspective, in the absence of a fitness cost, the migration threshold is 0; with a fitness cost of $s = 0.05$, the migration threshold is 0.016% per generation; and with a fitness cost of $s = 0.5$, the migration threshold is 2.8% per generation. Inter-village migration rates have been estimated at ~0.8% per generation for rural villages 7 km apart in Mali, West Africa (Taylor & Manoukis, 2003); although the initial sites for an open release of transgenic mosquitoes with gene drive systems are likely to be more isolated than this. We use 1% per generation as a conservative estimate of the migration rate between a release site and neighboring population.

With the above migration rates in mind, the source model predicts that *Medea* is very capable of spreading from one village to another in the presence of a small fitness cost. In the presence of a large fitness cost, *Medea* may be confined to its release site – e.g. for a fitness cost of $s = 0.5$ and a default migration rate (1% per generation), *Medea* allele frequency is only expected to reach 3.3% in population B – however, large fitness costs are unlikely to be relied upon for achieving confinement and are included here for illustrative purposes only. Catteruccia *et al.* (2003) estimate a fitness cost due to the insertion and/or expression of a transgene in *Anopheles* mosquito lines to be less than $s = 0.2$; however, fitness effects are expected to be transgene and insertion site-specific (Catteruccia *et al.*, 2003) and may also evolve over time, causing them to be a particularly unreliable mechanism for achieving confinement.

Two-population model

We now move onto a model in which migration is explicitly two-way, i.e. the mating pool in both populations is made up of individuals from both populations. For a migration rate of μ in both directions, this means that we need to make the following substitutions in the equations describing population A,

$$(u_{A,k}, v_{A,k}, w_{A,k}) \leftarrow (u_{A,k}, v_{A,k}, w_{A,k})(1 - \mu) + (u_{B,k}, v_{B,k}, w_{B,k})\mu, \quad (18)$$

Applying these substitutions to eqn (1), for illustrative purposes, we obtain,

$$\widehat{u}_{A,k+1} = [u_{A,k}(1 - \mu) + u_{B,k}\mu]^2 + 0.5[u_{A,k}(1 - \mu) + u_{B,k}\mu][v_{A,k}(1 - \mu) + v_{B,k}\mu]. \quad (21)$$

These substitutions also apply to eqns (2–3); while eqns (4–7) are unchanged. Analogous substitutions apply to population B. Considering a release in population A at generation 0, the initial condition for the difference equations is given by,

$$(u_{A,0}, w_{A,0}) = (1 - x, x), \quad (22)$$

$$(u_{B,0}, w_{B,0}) = (1, 0). \quad (23)$$

Here, the released mosquitoes represent a proportion, x , of population A at the time of release. Using this initial condition and the difference equations described above, we can calculate the time-series dynamics of the *Medea* allele in both populations through numerical iteration (analytic solutions are too complex to be useful).

A stochastic version of this model can be computed by calculating the expected genotype frequencies as previously described and then sampling a total of N individuals from a multinomial distribution parameterized by these frequencies (here, N represents the population size of reproducing adults). The multinomial sampling procedure is outlined in the Supplemental Material, §I. This approach is justified because insect species tend to produce a large number of offspring per adult; however the environment can only sustain a limited number of individuals (Service, 1993) leading to a situation in which individuals are essentially sampled from a larger population at each generation. Stochastic systems capture some of this randomness and, consequently, do not display clear threshold behavior – an allele may spread to fixation if it is released at sub-threshold levels according to a deterministic model; however, the probability that it fixes will tend to be low and depend on the population size, N . We consider a default population size of 10,000; but also consider smaller populations of 100 and 1,000 individuals.

Results

Fig. 1D–E depicts single stochastic realizations of the two-population model of *Medea* spread with a population size of 10,000. A construct with a fitness cost of $s = 0.05$ with heterozygosity $h = 0.5$ released in population A fixes in both populations (Fig. 1D); however, the same construct with a fitness cost of $s = 0.5$ only establishes itself in population A (Fig. 1E). The two-population and source models both display threshold behavior with respect to

migration rate; however, migration thresholds are slightly higher under the two-population model (Fig. 1F). This is due to the fact that, in the two-population model, prior to population replacement, incoming wild-type mosquitoes reduce the *Medea* allele frequency in population A, thus reducing the migration rate of mosquitoes having the *Medea* allele into population B. For the same reason, the *Medea* allele persists at slightly lower levels in population B under the two-population model (Fig. 1G). Although migration thresholds exist, they are unrealistically low for realistic fitness costs. This suggests that *Medea* is not confinable to its release site under these simple models.

It is worth noting that the migration thresholds in Fig. 1F correspond to deterministic versions of both the two-population and source models. Migration thresholds are less meaningful for stochastic versions of these models – for the source model, any migration will always lead to fixation in population B, eventually; and for the two-population model, the allele will ultimately be fixed or eliminated from both populations; however, the time taken to reach these absorbing states may be very long. For all practical purposes, migration thresholds do exist for stochastic models on a human timescale; however the stochastic thresholds are slightly lower because chance events make it easier for the *Medea* allele to cross the deterministic threshold above which it will be favored in population B.

We define the stochastic threshold for the two-population model as the migration rate above which the probability that the element becomes established in the neighboring population (i.e. spreads to a transgenic frequency above 90%) within 1,000 generations is greater than 0.1%. The stochastic threshold is dependent on population size, and we calculate this for populations having 100, 1,000 and 10,000 individuals. For the two-population model, a construct with a fitness cost of $s=0.5$ with heterozygosity $h=0.5$ has a deterministic migration threshold of 4.0% per generation; however this decreases to ~3.2% per generation for populations having 1,000 mosquitoes in each village and to ~0.9% per generation for a populations of 100 mosquitoes (Fig. 1H). This threshold difference is an important consideration to keep in mind for all of the gene drive systems analyzed in this paper since it highlights the potential for otherwise confineable systems to spread into small populations.

4. Other invasive gene drive systems

Medea is an invasive gene drive system in the sense that it is predicted to spread into neighboring populations unless it is associated with a very large fitness cost. Three other gene drive systems that fall into this category are TEs, HEGs and *Wolbachia*. We show that TEs are predicted to spread into neighboring populations whenever they are capable of spreading at all. HEGs are one of the most invasive of the gene drive systems being considered, but can be confined if they are associated with a large dominant fitness cost; and *Wolbachia* is predicted to be confineable for high fitness costs and/or low maternal transmission rates; but under realistic parameterizations, it is predicted to spread widely.

TEs

TEs are selfish genetic elements that are capable of transposing replicatively within the genome and hence spreading into a population from low initial frequencies (Charlesworth *et al.*, 1994; Fig. S1A). Their use as a gene drive system was inspired by the observation that *P* elements spread through most of the worldwide *Drosophila melanogaster* population within a few decades (Kidwell, 1983). A number of models have been proposed to model the spread of TEs through randomly-mating populations (Charlesworth & Charlesworth, 1983; Ribeiro & Kidwell, 1994; Marshall, 2008), all of which generally conclude that a TE will spread into a population provided that its transposition rate overcompensates for its associated fitness cost. The TE will then reach an equilibrium copy number in the population due to a number of mechanisms which slow down the rate of transposition with increasing

copy number (Townsend & Hartl, 2000). A TE may be lost in the early stages of spread (Le Rouzic & Capy, 2005; Marshall, 2009); however, this is unlikely to occur for a TE intentionally introduced at high frequency at its release site, with sustained migratory events introducing it into neighboring populations.

We extend an existing model of TE spread (Marshall, 2009) to account for migration from a source population and between two connected metapopulations (Supplemental Material, §II). The results suggest that, for default parameterizations (Table 1), the TE is expected to spread from one population to another, and to reach near-fixation in both populations within 100 generations (Fig. S1B). If the transposition rate is decreased (Fig. S1C) or the fitness cost is increased (Fig. S1D), the TE is still predicted to spread into both populations for any nonzero migration rate; however, if the fitness cost outweighs the transposition rate, the TE is expected to be lost from both populations. This model does not account for the tendency of TEs to jump locally (Tower *et al.*, 1993; Zhang *et al.*, 2003) and to home in on certain genomic regions (Guimond *et al.*, 2003); however this is not expected to change the overall dynamics other than by reducing the effective transposition rate (Rasgon & Gould, 2005). It also does not account for evolutionary considerations such as mutational inactivation (Struchiner *et al.*, 2009); however, these are less relevant on the timescale of a population replacement program. The general conclusion therefore holds that any TE capable of spreading is not confineable to an isolated population.

HEGs

HEGs, like TEs, are highly invasive genetic elements that are difficult to confine to isolated populations. HEGs spread by expressing an endonuclease which creates a double-stranded break on versions of the homologous chromosome lacking the HEG at the position where it occurs. Homologous DNA repair then copies the HEG to the cut chromosome (Fig. S1E), increasing its representation in subsequent generations. Strategies have been proposed that utilize this property for either population suppression or replacement (Burt, 2003), and recent advances have been made in the development of both (Windbichler *et al.*, 2008; 2011). Here, we consider the strategy of population replacement, which has been modeled by Deredec *et al.* (2008). The general conclusion is that a HEG will spread provided that its homing rate overcompensates for its fitness cost; however, for fitness costs that are partially or fully dominant, there is a range of parameter space for which the HEG displays threshold behavior – becoming either fixed or lost depending on its initial frequency. Like a TE, a HEG may be lost in the early stages of spread (Marshall, 2009); however, this is unlikely for a HEG intentionally introduced into the environment.

We extend the model of Deredec *et al.* (2008) to account for migration from a source population and between two connected metapopulations (Supplemental Material, §III). The results suggest that a HEG with additive fitness costs ($h = 0.5$) will spread from one population to another provided that the initial requirement for spread is satisfied, and the rate of spread will be significantly quicker than that for TEs (Fig. S1F). However, for partially or fully dominant fitness costs ($h > 0.5$), there are some regions of parameter space over which the HEG is predicted to be confineable to an isolated population (Fig. S1G–H). We include this observation for completeness; however it is unlikely to be of any practical use in terms of confinement because it requires that fitness costs fall within a very small range (e.g. $0.53 < s < 0.58$ for $h = 0.75$, $e = 0.5$ and $\mu = 0.01$), and fitness costs are both variable and notoriously difficult to measure under field conditions. This model does not account for the effects of gap repair with error-prone non-homologous end joining, which can create chromosomes with sequence alterations that make the chromosome resistant to further HEG invasion (Deredec *et al.* 2008). Furthermore, population suppression strategies utilizing HEGs display unique dynamics that must be considered independently (Burt, 2003).

Wolbachia

Wolbachia, a maternally-transmitted intracellular bacterium, is able to spread through a population by causing host reproductive alterations including cytoplasmic incompatibility (CI) (Stouthammer *et al.*, 1999), in which offspring of matings between infected males and uninfected females result in the death of some or all progeny, while matings involving infected females tend to produce infected offspring (Fig. S1I). This behavior biases the offspring ratio in favor of those carrying the *Wolbachia* infection, and *Wolbachia* is able to spread rapidly through a population despite a fitness cost (Turelli & Hoffmann, 1999). A recent environmental release of *Wolbachia*-infected *Aedes aegypti* in Australia has been successful in replacing the local wild population (Hoffmann *et al.*, 2011) and corresponding laboratory experiments suggest that the infected mosquitoes are unable to transmit dengue serotype 2 (Walker *et al.*, 2011). Like *Medea*, *Wolbachia* is predicted to spread into neighboring populations, but can be confined to its release site in the presence of inefficient transmission and/or a fitness cost (Flor *et al.*, 2007). The geographic spread of the *Wolbachia*-infected *Ae. aegypti* strain is currently being assessed.

A number of models have been proposed to describe the spread of *Wolbachia* through randomly-mating (Hoffmann *et al.*, 1990) and spatially-structured populations (Engelstadter & Telschow, 2009). Most relevant to this study, Flor *et al.* (2007) modeled *Wolbachia* in a two-population model for both mainland-island and two-way migration scenarios. Flor *et al.* (2007) concluded that infected and uninfected populations can co-exist provided that the migration rate between them is below a certain critical value; but that this value is very low for reasonable parameter estimates. Jansen *et al.* (2008) modeled the stochastic spread of *Wolbachia* in small populations beginning from sub-threshold frequencies and found its fixation probability to be significantly higher in small populations, sometimes exceeding that of a neutral allele. This suggests that, as for *Medea*, there is a chance that *Wolbachia* will spread between populations for migration rates smaller than the critical migration rates calculated by Flor *et al.* (2007).

We extend the model of Flor *et al.* (2007) to account for the stochasticity inherent in finite populations (Supplemental Material, §IV). The results suggest that, for default parameters (Table 1), the deterministic migration thresholds are unrealistically low (0.25% per generation), and are even lower for small populations (0.055% per generation for 1,000 individuals, and 0% per generation for 100 individuals) (Fig. S1K–L). A bacterium with a relatively large fitness cost ($s = 0.2$) is expected to be confineable to large populations (1,000–10,000 individuals) (Fig. S1J–K); but has a 47.5% chance of spreading between populations of 100 individuals. Life-shortening *Wolbachia* strains are being considered for the control of dengue fever-transmitting *Ae. aegypti* populations, and these strains are associated with large reductions in host fitness (Yeap *et al.*, 2010). However, it is not clear that a *Wolbachia*-induced fitness cost can be relied upon for confinement, since a *Wolbachia* infection in California *Drosophila simulans* populations has been observed to change over the course of 20 years from inducing a fecundity cost in females, to inducing a fecundity advantage (Weeks *et al.*, 2007).

Alternative modeling approaches confirm the ability of *Wolbachia* to spread despite a fitness cost. Turelli & Hoffmann (1991) modeled the spread of *Wolbachia* through the California *D. simulans* population as a “Bartonian wave” (Barton, 1979); and Schofield (2002) extended this model to incorporate rare long-range dispersal events which provide a better fit to the data. Wade & Stevens (1994) modeled the structure of a population by dividing it into a number of demes of finite size and modeling dispersal and competition at the level of the population with mating taking place within demes. This resulted in CI spread, but at a slower rate than in a panmictic population. Reuter *et al.* (2008), on the other hand, modeled competition at the local level, with a fraction of offspring dispersing from their natural

deme. This resulted in spread at a faster rate than in a panmictic population. Under all scenarios, the qualitative conclusion remains that *Wolbachia* is capable of spreading between populations; however, these divergent model predictions highlight the importance of having a good understanding of population structure in order to make reliable predictions of the rate and pattern of spatial spread.

It is worth noting that there is one exception to the above analysis – a *Wolbachia* infection is potentially confineable to an isolated release site if the neighboring population (and possibly the release site itself) is already infected with another *Wolbachia* strain. If the two strains cause matings between infected males harboring one *Wolbachia* strain and females harboring the other to be completely or partially unviable (because CI and rescue are brought about through different mechanisms in the two strains), the result is a bidirectional mating incompatibility termed bidirectional CI. Bidirectional CI displays threshold dynamics that depend on the properties of the two infections with respect to variables such as fitness cost, frequency of maternal transmission and efficiency of CI-induced sterility; however, for strains having similar properties, the more prevalent strain will tend to fix in the population (Keeling *et al.*, 2003). If a neighboring population is infected with the other strain, the two strains are predicted to stably coexist for migration rates greater than those observed in nature (Telschow *et al.*, 2005). Developing upon the model of Telschow *et al.* (2005), we find that a second *Wolbachia* infection displays very promising confinement properties under default parameters (Table 1), with a migration threshold of 13.3% per generation under the two-population model. The corresponding stochastic migration thresholds are 11.2% per generation for populations having 1,000 individuals and 7.0% per generation for populations having 100 individuals (Table 2). This suggests that a novel *Wolbachia* strain could be introduced into an isolated population in a confined manner if the surrounding population is already infected with another *Wolbachia* strain showing bidirectional CI with the introduced strain.

5. Underdominance

The simplest example of an underdominant system is a single bi-allelic locus for which the heterozygote is less fit than either homozygote (Hartl & Clark, 1997). In the extreme case, matings between opposite homozygotes are sterile, resulting in similar dynamics to bidirectional CI. Underdominant alleles therefore display high migration thresholds and are predicted to be confineable to partially-isolated populations (Altrock *et al.*, 2010). Attempts are currently underway to engineer single-allele underdominance (Hay *et al.*, unpublished); however, a range of alternative underdominant systems are also available, including chromosomal alterations such as inversions and translocations (Serebrovskii, 1940; Curtis, 1968), and underdominant chromosomes engineered using combinations of multiple toxins and antidotes at one or two loci (Davis *et al.*, 2001). We show that all these systems display desirable features for confinement, but differ in terms of their migration thresholds and the frequencies they reach in neighboring populations.

Translocations

Translocations are created when terminal segments of two nonhomologous chromosomes undergo a mutual exchange (Fig. 2A). These display underdominant dynamics because, when a translocation heterozygote undergoes meiosis, half its gametes lack one chromosomal segment and have a duplication of another, resulting in a reduction in the number of viable offspring. If the translocation-bearing gametes fuse with wild-type gametes, they produce unviable offspring; however, if they fuse with gametes having chromosomes with complementary duplications and deletions, they produce viable offspring (Fig. 2A). The use of translocations for transforming pest populations was initially suggested by Serebrovskii (1940) and later developed by Curtis (1968), and a number of

models have been proposed to describe their spread through randomly-mating populations (Wright, 1941; Curtis & Robinson, 1971). Models have also been proposed to describe the spatial spread of underdominant chromosomal rearrangements (Karlin & McGregor, 1972; Lande, 1979; Barton, 1979; Altrock *et al.*, 2010); however, these have addressed rearrangements that can be approximated by a single underdominant allele, while a proper treatment of reciprocal translocations requires at least two loci.

To address this discrepancy, we extend a model similar to that of Curtis & Robinson (1971) to describe the movement of individuals with reciprocal translocations between two connected metapopulations. We denote the first chromosome with a translocated segment by “*T*” and the wild-type version of this chromosome by “*t*.” Similarly, we denote the second chromosome with a translocated segment by “*R*” and the wild-type version of this chromosome by “*r*.” As a two-locus system, there are nine possible genotypes; however, only individuals carrying the full chromosome complement are viable, which corresponds to the genotypes *TTRR*, *TtRr* and *ttrr*. In population A, the proportion of the *k*th generation that are mosquitoes of genotypes *TTRR*, *TtRr* and *ttrr* are denoted by $p_{A,k}^{TTRR}$, $p_{A,k}^{TtRr}$ and $p_{A,k}^{ttrr}$, respectively. The corresponding proportions for population B are $p_{B,k}^{TTRR}$, $p_{B,k}^{TtRr}$ and $p_{B,k}^{ttrr}$.

We begin by considering the case without migration, assuming random mating and an infinite population size. The four haplotypes that determine genotype frequencies in the next generation – *TR*, *tR*, *Tr* and *tr* – are described by the following frequencies,

$$f_{A,k}^{TR} = p_{A,k}^{TTRR} + 0.25p_{A,k}^{TtRr}, \quad (24)$$

$$f_{A,k}^{tR} = 0.25p_{A,k}^{TtRr}, \quad (25)$$

$$f_{A,k}^{Tr} = 0.25p_{A,k}^{TtRr}, \quad (26)$$

$$f_{A,k}^{tr} = p_{A,k}^{ttrr} + 0.25p_{A,k}^{TtRr}. \quad (27)$$

By considering all possible mating pairs, the genotype frequencies in the next generation are given by,

$$p_{A,k+1}^{TTRR} = (f_{A,k}^{TR})^2 (1 - s) / W_{A,k+1}, \quad (28)$$

$$p_{A,k+1}^{TtRr} = 2(f_{A,k}^{TR} f_{A,k}^{tR} + f_{A,k}^{tR} f_{A,k}^{Tr}) (1 - hs) / W_{A,k+1}, \quad (29)$$

$$p_{A,k+1}^{ttrr} = (f_{A,k}^{tr})^2 / W_{A,k+1}. \quad (30)$$

Here, s represents the fitness cost of being homozygous for the translocation, hs represents the fitness cost on translocation heterozygotes, and $W_{A,k+1}$ is a normalizing term given by,

$$W_{A,k+1} = (f_{A,k}^{TR})^2 (1-s) + 2(f_{A,k}^{TR} f_{A,k}^{tr} + f_{A,k}^{tr} f_{A,k}^{Tr})(1-hs) + (f_{A,k}^{tr})^2. \quad (31)$$

For the two-population model, as for *Medea*, we assume that the mating pool in both populations is made up of individuals from both populations. This requires us to make the following substitutions,

$$(f_{A,k}^{TR}, f_{A,k}^{tr}, f_{A,k}^{Tr}, f_{A,k}^{tr}) \leftarrow (f_{A,k}^{TR}, f_{A,k}^{tr}, f_{A,k}^{Tr}, f_{A,k}^{tr})(1-\mu) + (f_{B,k}^{TR}, f_{B,k}^{tr}, f_{B,k}^{Tr}, f_{B,k}^{tr})\mu. \quad (32)$$

Applying these substitutions to eqn (29), for illustrative purposes, we obtain,

$$p_{A,k+1}^{TrR} = 2([f_{A,k}^{TR}(1-\mu) + f_{B,k}^{TR}\mu][f_{A,k}^{tr}(1-\mu) + f_{B,k}^{tr}\mu] + [f_{A,k}^{tr}(1-\mu) + f_{B,k}^{tr}\mu][f_{A,k}^{Tr}(1-\mu) + f_{B,k}^{Tr}\mu])(1-hs)/W_{A,k+1}, \quad (33)$$

and analogous substitutions apply to population B. Finally, considering a release in population A at generation 0, the initial condition for the difference equations is given by,

$$(p_{A,0}^{TTRR}, p_{A,0}^{trrr}) = (x, 1-x), \quad (34)$$

$$(p_{B,0}^{TTRR}, p_{B,0}^{trrr}) = (0, 1), \quad (35)$$

where the released mosquitoes represent a proportion, x , of population A at the time of release. The time-series dynamics of the translocations in both populations can then be calculated using the above difference equations.

Results from these simulations suggest that, for default migration rates (1% per generation) reciprocal translocations are expected to remain confined to partially-isolated populations. A translocation with a fitness cost of $s=0.05$ with heterozygosity $h=0.5$ is predicted to spread to near-fixation at its release site; but to reach a population frequency of only ~4.2% in neighboring populations (Fig. 2B). If the migration rate exceeds 5.8% per generation in the two-population model then, rather than spreading into population B, the translocation is expected to be lost from both populations (Fig. 2C); however migration rates at potential trial sites are expected to be lower than this. For populations of 1,000 individuals, there is a chance (greater than 0.1%) that the translocation will become established in both populations for migration rates higher than 6.0% per generation (Fig. 2D); and for populations of 100 individuals, the translocation has the same chance of becoming established in both populations for migration rates higher than 2.7% per generation. These

are lower migration thresholds than for bidirectional CI (Table 2); however they are still promising for the ability to confine a translocation to an isolated release site. For translocations with higher fitness costs, establishment in both populations is still possible for small populations (~100 individuals) with high migration rates (>5.9% per generation for $s=0.1$ and >19.2% per generation for $s=0.2$); but seems unlikely for larger populations (~1,000 individuals or larger). Translocations generated using x-rays often have high associated fitness costs, probably due to x-ray-induced background mutations, or the location of translocation breakpoints near genes with strong fitness effects (Robinson, 1976). This is thought to be a major reason why past field trials were unsuccessful in demonstrating spread (Lorimer *et al.*, 1972); however, it may now be possible to generate translocations with minimal fitness costs through the use of site-specific recombination systems that do not require mutagenesis (Golic & Golic, 1996; Egli *et al.*, 2004).

These results are qualitatively consistent with the single-locus models of underdominant chromosomal rearrangements mentioned earlier. Karlin & McGregor (1972) showed that spatial polymorphism – near-fixation in one population and near-elimination in another – is locally stable for sufficiently small migration rates. Altrock *et al.* (2010) extended this analysis, investigating the effects of asymmetries in fitnesses and migration rates, and confirmed that underdominant alleles can be used to achieve local population replacement. Barton (1979) modeled underdominant chromosomal rearrangements in continuous populations and showed that one arrangement can advance or recede if it has a higher or lower fitness than the other; however, this spread can be arrested in regions of low population density. Finally, it is worth noting that translocations are known to have accumulated and spread between populations in many species, presumably beginning from very low frequencies (Bush *et al.*, 1977); however, such events have occurred on an evolutionary timescale involving many natural trials (Wright, 1941), while the scenario being considered here involves a much shorter timescale and many fewer trials. Under these circumstances, reciprocal translocations with a modest fitness cost are expected to be confineable to a partially-isolated population.

Engineered underdominance (two-locus)

A novel form of underdominance that is in principle straightforward to engineer has been suggested by Davis *et al.* (2001). In this system there are two transgenic constructs, each of which possesses a gene whose expression induces lethality and a gene that suppresses the expression or activity of the gene inducing lethality carried by the other construct. The constructs can either be inserted at the same locus on a pair of homologous chromosomes (Fig. 2I) or at different loci on nonhomologous chromosomes (Fig. 2E). These systems display underdominant properties because individuals carrying neither or both constructs are viable; but a proportion of their offspring – those carrying just one of the constructs – are unviable. The dynamics of engineered underdominance have been modeled by Davis *et al.* (2001) and expanded upon by Magori & Gould (2006); however both of these modeling efforts have been limited to single randomly-mating, deterministic populations. We extend the model of Davis *et al.* (2001) to describe the dynamics of engineered underdominance in two connected metapopulations, beginning with the two-locus case (Fig. 2E).

Following from the model of reciprocal translocations, we denote the transgenic allele at the first locus by “*T*” and the null allele at this locus by “*t*.” At the second locus, we denote the transgenic allele by “*R*” and the null allele by “*r*.” As a two-locus system, there are nine possible genotypes; however, individuals possessing only one transgenic construct express a lethal gene without its suppressor, and so only five of these genotypes are viable – *TTRR*, *TtRR*, *TTRr*, *TtRr* and *ttrr*. In population A, we denote the proportion of the k th generation that are mosquitoes having these genotypes by $p_{A,k}^{TTRR}$, $p_{A,k}^{TtRR}$, $p_{A,k}^{TTRr}$, $p_{A,k}^{TtRr}$ and $p_{A,k}^{ttrr}$,

respectively, with corresponding symbols for population B. The four haplotypes that determine genotype frequencies in the next generation are TR , tR , Tr and tr , which are described by the following frequencies,

$$f_{A,k}^{TR} = p_{A,k}^{TTRR} + 0.5p_{A,k}^{TiRR} + 0.5p_{A,k}^{TTRr} + 0.25p_{A,k}^{TiRr}, \quad (36)$$

$$f_{A,k}^{tR} = 0.5p_{A,k}^{TiRR} + 0.25p_{A,k}^{TiRr}, \quad (37)$$

$$f_{A,k}^{Tr} = 0.5p_{A,k}^{TTRr} + 0.25p_{A,k}^{TiRr}, \quad (38)$$

$$f_{A,k}^{tr} = p_{A,k}^{ttrr} + 0.25p_{A,k}^{TiRr}. \quad (39)$$

By considering all possible mating pairs (assuming random mating and an infinite population size), the genotype frequencies in the next generation are given by,

$$p_{A,k+1}^{TTRR} = (f_{A,k}^{TR})^2 (1 - 2s) / W_{A,k+1}, \quad (40)$$

$$p_{A,k+1}^{TiRR} = 2f_{A,k}^{TR} f_{A,k}^{tR} (1 - 1.5s) / W_{A,k+1}, \quad (41)$$

$$p_{A,k+1}^{TTRr} = 2f_{A,k}^{TR} f_{A,k}^{Tr} (1 - 1.5s) / W_{A,k+1}, \quad (42)$$

$$p_{A,k+1}^{TiRr} = 2(f_{A,k}^{TR} f_{A,k}^{tr} + f_{A,k}^{tR} f_{A,k}^{Tr}) (1 - s) / W_{A,k+1}, \quad (43)$$

$$p_{A,k+1}^{ttrr} = (f_{A,k}^{tr})^2 / W_{A,k+1}. \quad (44)$$

Here, each construct is associated with a fitness cost of $s/2$ (i.e. $h=0.5$), and $W_{A,k+1}$ is a normalizing term given by,

$$W_{A,k+1} = (f_{A,k}^{TR})^2 (1 - 2s) + 2(f_{A,k}^{TR} f_{A,k}^{tR} + f_{A,k}^{TR} f_{A,k}^{Tr}) (1 - 1.5s) + 2(f_{A,k}^{TR} f_{A,k}^{tr} + f_{A,k}^{tR} f_{A,k}^{Tr}) (1 - s) + (f_{A,k}^{tr})^2. \quad (45)$$

Analogous equations apply to population B.

For the two-population model, we assume that the mating pool in both populations is made up of individuals from both populations, which requires us to make the substitutions in eqn (32). For illustrative purposes, applying these substitutions to eqn (43), we obtain,

$$p_{A,k+1}^{TtRr} = 2([f_{A,k}^{TR}(1-\mu) + f_{B,k}^{TR}\mu][f_{A,k}^{tr}(1-\mu) + f_{B,k}^{tr}\mu] + [f_{A,k}^{tR}(1-\mu) + f_{B,k}^{tR}\mu][f_{A,k}^{Tr}(1-\mu) + f_{B,k}^{Tr}\mu])(1-s)/W_{A,k+1}. \quad (46)$$

These substitutions also apply to eqns (40–44), and analogous substitutions apply to population B. Considering a release in population A at generation 0, the initial condition for the difference equations is given by,

$$(p_{A,0}^{TTRR}, p_{A,0}^{ttrr}) = (x, 1 - x), \quad (47)$$

$$(p_{B,0}^{TTRR}, p_{B,0}^{ttrr}) = (0, 1), \quad (48)$$

where the released mosquitoes represent a proportion, x , of population A at the time of release. The time-series dynamics in both populations can then be calculated using the above difference equations.

Numerical simulations suggest that, for default migration rates (1% per generation) two-locus engineered underdominance is expected to remain confined to partially-isolated populations, although it is more likely to spread between populations than reciprocal translocations. Fig. 2F depicts a single stochastic realization of the two-population model for constructs having additive fitness costs of $s = 0.05$. For a migration rate of 1% per generation, the system spreads to near-fixation at its release site, but only reaches a frequency of ~3.2% in neighboring populations. For the same fitness cost, the migration threshold under the two-population model is 4.3% per generation (Fig. 2G), above which the system becomes established in both populations. As population size decreases, the stochastic migration threshold also decreases, falling to 3.7% per generation for populations of 1,000 individuals, and 1.5% per generation for populations of 100 individuals (Fig. 2H). This is lower than for translocations, which are more likely to be lost from both populations at higher migration rates; however two-locus engineered underdominance has a significantly lower release threshold (33.3% c.f. 52.8% for translocations for $s = 0.05$), meaning that fewer transgenic insects need to be released for local population replacement to be achieved. Furthermore, for sub-threshold migration rates, two-locus engineered underdominance spreads to lower frequencies in neighboring populations (Table 2).

Engineered underdominance (single-locus)

The two-construct underdominance system proposed by Davis *et al.* (2001) also works if both constructs are inserted at a single locus, although with different dynamical properties including higher release and migration thresholds. As for the two-locus system, there are two transgenic constructs, each possessing a lethal gene and suppressor for the lethal gene on the other construct (Fig. 2I).

Here, we extend the model of Davis *et al.* (2001) to describe the dynamics of single-locus engineered underdominance under the source and two-population models. As a three-allele system – two transgenic alleles, “*T*” and “*R*,” and one null allele, “*t*” – there are six possible genotypes; however, since individuals possessing only one transgenic construct express a lethal gene without its suppressor, only two of these genotypes are viable – *tt* and *TR*. In population A, we denote the proportion of the *k* th generation that are mosquitoes having these genotypes by $u_{A,k}$ and $w_{A,k}$, respectively, with corresponding symbols for population B. The genotype frequencies in the next generation are then described by the following frequencies,

$$u_{A,k+1} = u_{A,k}^2 / (u_{A,k}^2 + 0.5w_{A,k}^2 (1 - s)), \quad (49)$$

$$w_{A,k+1} = 0.5w_{A,k}^2 (1 - s) / (u_{A,k}^2 + 0.5w_{A,k}^2 (1 - s)). \quad (50)$$

Analogous equations apply to population B.

For the source model, we begin by assuming that the transgene has already reached equilibrium in population A, ignoring migratory effects. Davis *et al.* (2001) showed that, in a randomly-mating population, single-locus engineered underdominance spreads to fixation for releases exceeding 66.7% in the absence of a fitness cost, and 67.8% in the presence of a 5% fitness cost. We assume a super-threshold release in population A, and that population A remains fixed for this genotype, donating transgenic individuals to population B at a rate, μ , measured relative to the size of population B. The genotype frequencies in population B are then given by,

$$u_{B,k+1} = u_{B,k}^2 / (u_{B,k}^2 + 0.5w_{B,k}^2 (1 - s) + \mu), \quad (51)$$

$$w_{B,k+1} = (0.5w_{B,k}^2 (1 - s) + \mu) / (u_{B,k}^2 + 0.5w_{B,k}^2 (1 - s) + \mu). \quad (52)$$

To characterize the dynamics of the introduced genotype in population B, we begin by calculating the equilibria,

$$w_{B,k+1} = w_{B,k} = w_{B,*}. \quad (53)$$

This gives us three solutions,

$$w_{B,*} = \left\{ 1, \frac{1 - \sqrt{1 - 2\mu(3 - s)}}{3 - s}, \frac{1 + \sqrt{1 - 2\mu(3 - s)}}{3 - s} \right\}. \quad (54)$$

The first of these represents fixation, and the second and third represent a mixture of homozygotes and wild-types. Calculating the stabilities of these equilibria, we see that

fixation is stable, and the other two equilibria only exist when $\mu \leq 1/(6 - 2s)$. If this condition is satisfied, then the second equilibrium is stable and the third is unstable. This implies that,

$$w_{B,s} = \begin{cases} \frac{1 - \sqrt{1 - 2\mu(3-s)}}{3-s}, & \mu \leq \frac{1}{6-2s} \\ 1, & \mu > \frac{1}{6-2s} \end{cases} \quad (55)$$

The expression, $1/(6 - 2s)$, therefore represents a migration threshold for the source model, above which the introduced allele becomes fixed in population B, and below which it merely persists at a low level. This migration threshold is 16.7% per generation in the absence of a fitness cost, and 16.9% per generation for a fitness cost of $s = 0.05$ – much greater than the default migration rate (1% per generation) – suggesting that single-locus engineered underdominance will remain confined to its release site under all realistic parameter values.

A corresponding two-population model confirms that single-locus engineered underdominance is confineable to a partially-isolated population (Supplementary Material, §V) and that, rather than spreading into neighboring populations at high migration rates, the system is eliminated from both populations, similar to the case for reciprocal translocations. For a fitness cost of $s = 0.05$, the system is predicted to spread to near-fixation at its release site; but to reach a population frequency of only ~0.005% in neighboring populations (Fig. 2J). The migration threshold leading to loss is much higher than that for translocations (9.7% per generation c.f. 5.8% per generation for translocations, $s = 0.05$) (Fig. 2K); however, it is also the case that the release threshold is much higher (67.8% c.f. 52.8% for translocations, $s = 0.05$), indicating that many more transgenic insects need to be released to exceed the threshold required for local population replacement. Interestingly, when population sizes are reduced to 100 individuals, single-locus engineered underdominance is still not expected to become established in neighboring populations (Fig. 2L). The single-locus system is therefore highly confineable to isolated populations, and its main disadvantage is the high release requirement.

Underdominance (single-allele)

The confinement of underdominant alleles to local populations has been modeled by Karlin & McGregor (1972) and Altrock *et al.* (2010). These models apply to a single bi-allelic locus with reduced heterozygote fitness, and may be adequate for describing chromosomal rearrangements such as paracentric inversions and fusions (Lande, 1979). Underdominant alleles display release thresholds similar to those for translocations; however their confinement properties tend to be stronger (Altrock *et al.*, 2010). That said; the strength of confinement is predicted to decline with increasing heterozygote fitness – a situation which has been described for paracentric inversions and fusions (Lande, 1979). We are specifically interested in the extreme case of underdominance, in which heterozygotes are completely unviable (Fig. 2M), and molecular techniques are currently being investigated to engineer these systems (Hay *et al.*, unpublished).

We use a simplified version of the models of Karlin & McGregor (1972) and Altrock *et al.* (2010) to describe the dynamics of an extreme underdominant allele under the source and two-population models (Supplementary Material, §VI). The source model is analytically tractable because there is only one parameter – the fitness cost of being homozygous for the introduced allele, s . Following a similar procedure to the case of single-locus engineered underdominance, we derive a migration threshold of $1/(8 - 4s)$ for the source model, which equates to 12.5% in the absence of a fitness cost, and 13.2% per generation for a fitness cost of $s = 0.05$. This is much greater than the default rate of 1% per generation, suggesting that

single-locus engineered underdominance will remain confined to its release site under all realistic parameter values.

Analysis of the corresponding two-population model confirms that an extreme underdominant allele is confineable to a partially-isolated population (Fig. 2N) and that, like single-locus engineered underdominance and reciprocal translocations, the system is eliminated from both populations at high migration rates (Fig. 2O). For populations of 1,000 individuals, there is a chance (greater than 0.1%) that the underdominant allele will become established in both populations for migration rates higher than 19.4% per generation (Fig. 2P); and for populations of 100 individuals, the underdominant allele has the same chance of becoming established in both populations for migration rates higher than 14.7% per generation (both of these migration rates are much higher than those expected in nature). Interestingly, while translocations and extreme underdominant alleles have very similar release thresholds (51.3% for underdominant alleles c.f. 52.8% for translocations, $s=0.05$), the migration threshold leading to loss is much higher for an extreme underdominant allele (17.6% per generation c.f. 5.8% for translocations, $s=0.05$) (Fig. 2O), and these alleles are predicted to spread to much lower levels in neighboring populations (~0.01% c.f. ~4.2% for translocations). An extreme underdominant allele is therefore preferable to a pair of reciprocal translocations for local replacement since it is predicted to lead to less contamination of neighboring populations and be less vulnerable to elimination due to inward migration.

6. Other non-invasive gene drive systems

The four forms of underdominance outlined above are non-invasive in the sense that they are predicted to spread at their release site but only persist at low levels in neighboring populations. We show that three other gene drive systems belong to this category – *Semele*, inverse *Medea* and *Merea*. These systems each manipulate the offspring ratio in different ways by favoring one allele over another through the targeted action of a toxin and antidote encoded at a single locus

Semele

Semele consists of a toxin expressed in the semen of transgenic males and an antidote expressed in females (Fig. 3A). An all-male release results in population suppression because wild females that mate with transgenic males produce no offspring; and a release that includes transgenic females results in gene drive because females having the construct are favored at high population frequencies (Marshall *et al.*, 2011). The dynamics of *Semele* in a single, randomly-mating population have been modeled by Marshall *et al.* (2011), with particular emphasis on its potential application for localized population replacement. Here, we extend this model to describe the dynamics in two connected metapopulations.

We consider a *Semele* element as a single allele which we denote by “*T*” and denote the null allele at this locus by “*t*.” In population A, the proportion of the k th generation that are mosquitoes of genotypes tt , Tt and TT are denoted by $u_{A,k}$, $v_{A,k}$ and $w_{A,k}$, respectively. The corresponding proportions for population B are $u_{B,k}$, $v_{B,k}$ and $w_{B,k}$. We begin by considering the case without migration, assuming random mating and an infinite population size. By considering all possible mating pairs, and removing unproductive crosses between transgenic males and tt females, the genotypes of embryos in the next generation in population A are described by the ratio $\hat{u}_{A,k+1} : \hat{v}_{A,k+1} : \hat{w}_{A,k+1}$, where,

$$\hat{u}_{A,k+1} = u_{A,k}^2 + 0.25v_{A,k}^2 + 0.5u_{A,k}v_{A,k}, \quad (56)$$

$$\widehat{v}_{A,k+1} = u_{A,k} w_{A,k} + 0.5 u_{A,k} v_{A,k} + 0.5 v_{A,k}^2 + w_{A,k} v_{A,k}, \quad (57)$$

$$\widehat{w}_{A,k+1} = w_{A,k}^2 + w_{A,k} v_{A,k} + 0.25 v_{A,k}^2. \quad (58)$$

Here, we have assumed 100% toxin efficiency. The genotype frequencies in the next generation are given by,

$$u_{A,k+1} = \widehat{u}_{A,k+1} / W_{A,k+1}, \quad (59)$$

$$v_{A,k+1} = \widehat{v}_{A,k+1} (1 - hs) / W_{A,k+1}, \quad (60)$$

$$w_{A,k+1} = \widehat{w}_{A,k+1} (1 - s) / W_{A,k+1}. \quad (61)$$

Here, s and hs represent the fitness costs associated with being homozygous or heterozygous for the *Semele* element, and $W_{A,k+1}$ is a normalizing term given by,

$$W_{A,k+1} = \widehat{u}_{A,k+1} + \widehat{v}_{A,k+1} (1 - hs) + \widehat{w}_{A,k+1} (1 - s). \quad (62)$$

Analogous equations apply to population B.

For the two-population model, we assume that the mating pool in both populations is made up of individuals from both populations, which requires us to make the same substitutions as for *Medea*, i.e. eqn (18). Applying these substitutions to eqn (57), for illustrative purposes, we obtain,

$$\begin{aligned} \widehat{v}_{A,k+1} = & [u_{A,k} (1 - \mu) + u_{B,k} \mu] [w_{A,k} (1 - \mu) + w_{B,k} \mu] \\ & + 0.5 [v_{A,k} (1 - \mu) + v_{B,k} \mu]^2 + 0.5 [u_{A,k} (1 - \mu) + u_{B,k} \mu] [v_{A,k} (1 - \mu) + v_{B,k} \mu] \\ & + [w_{A,k} (1 - \mu) + w_{B,k} \mu] [v_{A,k} (1 - \mu) + v_{B,k} \mu]. \end{aligned} \quad (63)$$

These substitutions apply to eqns (56–58), and analogous substitutions also apply to population B. Considering a release in population A at generation 0, the initial condition for the difference equations is given by,

$$(u_{A,0}, w_{A,0}) = (1 - x, x), \quad (64)$$

$$(u_{A,0}, w_{A,0}) = (1, 0), \quad (65)$$

where the released mosquitoes represent a proportion, x , of population A at the time of release. The time-series dynamics of the *Semele* element can then be calculated for both populations.

Results from these simulations suggest that, for default migration rates (1% per generation), *Semele* can be confined to a partially-isolated population and that, in general, *Semele* displays an intermediate strength of confinement, somewhere between that of translocations and single-allele underdominance. For a fitness cost of $s = 0.05$ with heterozygosity $h = 0.5$, the element is predicted to spread to near-fixation at its release site, and to reach a population frequency of $\sim 1.8\%$ in neighboring populations (Fig. 3B, c.f. $\sim 0.01\%$ for single-allele underdominance and $\sim 4.2\%$ for translocations). The migration threshold for *Semele* is also intermediate (6.9% c.f. 17.6% for single-allele underdominance and 5.8% for translocations, Fig. 3C), despite its release threshold being lower than that for either underdominant system (38.9% c.f. 51.3% for single-allele underdominance and 52.8% for translocations, $s = 0.05$, $h = 0.5$). For smaller population sizes, there is a chance that *Semele* will spread to neighboring populations for smaller migration rates (Fig. 3D); however these are still significantly higher than the migration rates expected between trial sites. This suggests that *Semele* is a reasonable choice for local population replacement due to its predicted confinability and moderate release size required for replacement. The *Semele* system has not yet been engineered; although several possible approaches are outlined by Marshall *et al.* (2011), and since these only involve a single toxin-antidote pair, they may be simpler to construct than the double toxin-antidote pair systems proposed by Davis *et al.* (2001).

Inverse Medea

Inverse *Medea* is another single locus toxin-antidote system that displays threshold behavior. It consists of a zygotic toxin and maternal antidote (Fig. 3E) which render heterozygous offspring of wild-type mothers unviable (Marshall & Hay, 2011a). This enables it to spread when it represents a majority of the alleles in a population. The potential for inverse *Medea* to be confined to a partially-isolated population has been demonstrated by Marshall & Hay (2011a), and we expand on their analysis here through the inclusion of a stochastic implementation.

Simulations suggest that, while inverse *Medea* can be confined to a partially-isolated population, the system is particularly vulnerable to elimination due to inward migration. For a fitness cost of $s = 0.05$ with heterozygosity $h = 0.5$, the element is predicted to spread to a high frequency at its release site, and to reach a population frequency of $\sim 1.6\%$ in neighboring populations (Fig. 3F). For the same fitness cost, the migration threshold leading to loss is only 1.7% per generation (Fig. 3G), although this increases to 2.1% per generation if the fitness cost is dominant (Marshall & Hay, 2011a). For small population sizes (~ 100 individuals), there is a chance that inverse *Medea* will become established in neighboring populations; however this is only expected to occur for migration rates above 6% per generation (Fig. 3H). This suggests that, although inverse *Medea* is confineable – it minimally contaminates neighboring populations and is eliminated at high migration rates – it is not the best candidate for local replacement since it is easily eliminated from its release site. This is less of a concern if the release site has a relatively large population size compared to neighboring locations; but is of greater concern when mosquito population structure is not well characterized. The inverse *Medea* system has not yet been engineered; although several possible approaches are outlined by Marshall & Hay (2011a).

Merea

Merea is a variant of the standard *Medea* element described earlier in which the antidote is recessive (requiring inheritance of copies from both parents) rather than dominant (Fig. 3I). Heterozygous offspring are no longer rescued by one copy of the zygotic antidote and therefore the element must exceed a certain critical frequency before the offspring ratio is distorted in its favor. The dynamics of *Merea* in a single randomly-mating population have been modeled by Marshall & Hay (2011b) as part of a larger analysis of single-construct toxin-antidote systems. Here, we extend the analysis to describe the stochastic and deterministic dynamics in two connected metapopulations.

We consider the *Merea* element as a single allele which we denote by “*M*,” and refer to the null allele as “*m*.” In population A, the proportion of the *k* th generation that are mosquitoes of genotypes *mm*, *Mm* and *MM* are denoted by $u_{A,k}$, $v_{A,k}$ and $w_{A,k}$, respectively, with corresponding symbols for population B. We begin by considering the case without migration, assuming random mating and an infinite population size. By considering all possible mating pairs, the genotypes of embryos in the next generation in population A are described by the ratio $\hat{u}_{A,k+1} : \hat{v}_{A,k+1} : \hat{w}_{A,k+1}$, where,

$$\hat{u}_{A,k+1} = u_{A,k}^2 + 0.5u_{A,k}v_{A,k}, \quad (66)$$

$$\hat{v}_{A,k+1} = u_{A,k}w_{A,k} + 0.5u_{A,k}v_{A,k}, \quad (67)$$

$$\hat{w}_{A,k+1} = w_{A,k}^2 + w_{A,k}v_{A,k} + 0.25v_{A,k}^2. \quad (68)$$

Here, we have assumed 100% toxin efficiency (Chen *et al.*, 2007). The rest of the analysis is identical to that for *Semele* and inverse *Medea*, since all three are single locus toxin-antidote systems, and is described in eqns (59–65). Once again, the equations for the two-population model can be iterated to calculate the time-series dynamics of the *Merea* element in both populations.

Simulations suggest that, for default migration rates (1% per generation), *Merea* can be confined to a partially-isolated population and that, like *Semele*, *Merea* displays an intermediate strength of confinement between that of translocations and single-allele underdominance. For a fitness cost of $s = 0.05$ with heterozygosity $h = 0.5$, the element is predicted to spread to near-fixation at its release site, and to reach a population frequency of ~2.0% in neighboring populations (Fig. 3J; c.f. ~1.8% for *Semele*). It has a migration threshold of 7.0% per generation (Fig. 3K, c.f. 6.9% per generation for *Semele*) and a release threshold of 42.8% (c.f. 38.9% for *Semele*, $s = 0.05$, $h = 0.5$). For smaller population sizes, there is a chance that *Merea* will spread to neighboring populations for smaller migration rates (Fig. 3L); however, for populations of 100 individuals, the stochastic migration threshold is ~4.0% (Table 2), which is still significantly higher than migration rates expected between trial sites. *Merea* performs very similarly to *Semele* in terms of confinement and would be a reasonable choice for local population replacement. An advantage of *Merea* is its speed and efficiency of spread – it takes only seven generations to exceed a frequency of 98% in population A following a 50% release, c.f. 36 generations for *Semele* ($s = 0.05$, $h = 0.5$, $\mu = 0.01$ per generation). It is also capable of spreading to fixation in an isolated population despite a fitness cost, while *Semele* is not. However, the *Merea* system has not

yet been engineered and does require engineering a recessive antidote, which has not yet been achieved.

7. Killer-rescue

A novel two-locus gene drive system has been suggested by Gould *et al.* (2008) that is self-limiting in time. The system consists of two alleles at unlinked loci – one that encodes a toxin (a killer allele), and another that confers immunity to the toxin (a rescue allele), which could be linked to a gene for disease refractoriness. A release of individuals homozygous for both alleles initially results in drive as the alleles segregate and the presence of the killer allele in the population confers a temporary benefit to those also carrying the rescue allele; however, as the killer allele declines in frequency, the rescue allele is eventually eliminated through natural selection if it confers a fitness cost to carriers. The dynamics of the killer-rescue system are distinct in the sense that gene drive is temporary, and the degree to which transgenic alleles spread depends on their release size and fitness costs. The dynamics are also distinct in the sense that killer-rescue does not display threshold behavior and, as we show, transgenic alleles are therefore able to disperse into neighboring populations with comparable ease, albeit over a limited time span. Gould *et al.* (2008) have analyzed the dynamics of the killer-rescue system in a single, randomly-mating population, and here, we extend this analysis to two connected metapopulations.

As a two-locus system, there are nine possible genotypes for killer-rescue; however only seven of these are viable (individuals possessing the killer construct without the rescue construct are unviable). We denote the killer allele at the first locus by “*K*” and the null allele at this locus by “*k*.” At the second locus, we denote the rescue allele by “*R*” and the null allele by “*r*.” In population A, the proportion of the *k*th generation that are mosquitoes of genotypes *KKRR*, *KKRr*, *KkRR*, *KkRr*, *kkRR*, *kkRr* and *kkrr* are denoted by

$p_{A,k}^{KKRR}$, $p_{A,k}^{KKRr}$, $p_{A,k}^{KkRR}$, $p_{A,k}^{KkRr}$, $p_{A,k}^{kkRR}$, $p_{A,k}^{kkRr}$ and $p_{A,k}^{kkrr}$, respectively. The corresponding proportions for population B are $p_{B,k}^{KKRR}$, $p_{B,k}^{KKRr}$, $p_{B,k}^{KkRR}$, $p_{B,k}^{KkRr}$, $p_{B,k}^{kkRR}$, $p_{B,k}^{kkRr}$ and $p_{B,k}^{kkrr}$.

We begin by considering the case without migration, assuming random mating and an infinite population size. The four haplotypes that determine genotype frequencies in the next generation are *KR*, *Kr*, *kR* and *kr*. These are described by the following frequencies,

$$f_{A,k}^{KR} = p_{A,k}^{KKRR} + 0.5p_{A,k}^{KKRr} + 0.5p_{A,k}^{KkRR} + 0.25p_{A,k}^{KkRr}, \quad (69)$$

$$f_{A,k}^{Kr} = 0.5p_{A,k}^{KKRr} + 0.25p_{A,k}^{KkRr}, \quad (70)$$

$$f_{A,k}^{kR} = p_{A,k}^{kkRR} + 0.5p_{A,k}^{kkRr} + 0.5p_{A,k}^{KkRR} + 0.25p_{A,k}^{KkRr}, \quad (71)$$

$$f_{A,k}^{kr} = p_{A,k}^{kkrr} + 0.5p_{A,k}^{kkRr} + 0.25p_{A,k}^{KkRr}. \quad (72)$$

By considering all possible mating pairs, the genotype frequencies in the next generation are given by,

$$p_{A,k+1}^{KKRR} = (f_{A,k}^{KR})^2 (1 - 2s) / W_{A,k+1}, \quad (73)$$

$$p_{A,k+1}^{KKRr} = 2f_{A,k}^{KR} f_{A,k}^{Kr} (1 - 1.5s) / W_{A,k+1}, \quad (74)$$

$$p_{A,k+1}^{KkRR} = 2f_{A,k}^{KR} f_{A,k}^{kR} (1 - 1.5s) / W_{A,k+1}, \quad (75)$$

$$p_{A,k+1}^{KkRr} = 2(f_{A,k}^{KR} f_{A,k}^{kr} + f_{A,k}^{kR} f_{A,k}^{Kr}) (1 - s) / W_{A,k+1}, \quad (76)$$

$$p_{A,k+1}^{kkRR} = (f_{A,k}^{kR})^2 (1 - s) / W_{A,k+1}, \quad (77)$$

$$p_{A,k+1}^{kkRr} = 2f_{A,k}^{kr} f_{A,k}^{kR} (1 - 0.5s) / W_{A,k+1}, \quad (78)$$

$$p_{A,k+1}^{kkrr} = (f_{A,k}^{kr})^2 / W_{A,k+1}. \quad (79)$$

Here, each construct is associated with a fitness cost of $s/2$ (i.e. $h=0.5$), and $W_{A,k+1}$ is a normalizing term given by,

$$W_{A,k+1} = (f_{A,k}^{KR})^2 (1-2s) + (2f_{A,k}^{KR} f_{A,k}^{Kr} + 2f_{A,k}^{KR} f_{A,k}^{kR}) (1-1.5s) + ((f_{A,k}^{kR})^2 + 2f_{A,k}^{KR} f_{A,k}^{kr} + 2f_{A,k}^{Kr} f_{A,k}^{kR}) (1-s) + 2f_{A,k}^{kr} f_{A,k}^{kR} (1-0.5s) + (f_{A,k}^{kr})^2. \quad (80)$$

Analogous equations apply to population B.

The source model in which the system first reaches equilibrium in population A is not relevant because the killer-rescue system is designed to be self-limiting in time, and hence the only equilibrium is loss. We therefore consider the two-population model to get an idea of how much the killer-rescue system spreads into neighboring populations. For a migration rate of μ in both directions, we make the following substitutions,

$$(f_{A,k}^{KR}, f_{A,k}^{kR}, f_{A,k}^{Kr}, f_{A,k}^{kr}) \leftarrow (f_{A,k}^{KR}, f_{A,k}^{kR}, f_{A,k}^{Kr}, f_{A,k}^{kr}) (1 - \mu) + (f_{B,k}^{KR}, f_{B,k}^{kR}, f_{B,k}^{Kr}, f_{B,k}^{kr}) \mu, \quad (81)$$

Applying these substitutions to eqn (76), for illustrative purposes, we obtain,

$$p_{A,k+1}^{KkRr} = 2([f_{A,k}^{KR}(1-\mu) + f_{B,k}^{KR}\mu][f_{A,k}^{kr}(1-\mu) + f_{B,k}^{kr}\mu] + [f_{A,k}^{kr}(1-\mu) + f_{B,k}^{kr}\mu][f_{A,k}^{KR}(1-\mu) + f_{B,k}^{KR}\mu])(1-s)/W_{A,k+1}. \quad (82)$$

These substitutions also apply to eqns (73–80) and analogous substitutions apply to population B. Considering a release in population A at generation 0, the initial condition for the difference equations is given by,

$$(p_{A,0}^{kkr}, p_{A,0}^{KKRR}) = (1-x, x), \quad (83)$$

$$(p_{B,0}^{kkr}, p_{B,0}^{KKRR}) = (1, 0). \quad (84)$$

Here, the released mosquitoes represent a proportion, x , of population A at the time of release. Using this initial condition and the difference equations described above, we can calculate the time-series dynamics of the killer-rescue alleles in both populations.

We consider two realizations of killer-rescue dynamics with default migration rates (1% per generation). Fig. 4A depicts a single stochastic realization of a 50% release in population A of individuals homozygous for both alleles, each associated with a fitness costs of $s=0.05$. Transgenic mosquitoes reach a peak frequency of 95% in population A and 29% in population B, and these frequencies decline slowly over the next 200 generations. Fig. 4B depicts a release every time transgenic frequency falls below 50%. Transgenic mosquitoes reach a peak frequency of 98% in population A and 43% in population B, which oscillates between this and 31%.

Killer-rescue can be more effectively confined with higher fitness costs and smaller release sizes. The catch-22 is that, with higher fitness costs and/or smaller release sizes, transgenic mosquitoes also reach a lower maximum frequency at their release site. Fig. 4C depicts the multiple release strategy mentioned above for alleles having a fitness cost of $s=0.1$. For a single release, transgenic mosquitoes reach a peak frequency of 89% in population A and 16% in population B. For multiple releases, transgenic mosquitoes reach a peak frequency of 94% in population A and 23% in population B, oscillating between this and 16%. The same correlation between reduced transgenic frequency at the release site and reduced contamination of neighboring populations is seen when the release size is reduced. Thus, while killer-rescue is self-limiting in time (assuming that the fitness load on the rescue allele persists), it spreads to higher frequencies in neighboring populations in the short term than gene drive systems with thresholds.

8. Spread beyond immediately neighboring populations

Finally, to monitor the spread of a gene drive system beyond its neighboring population, we adapt the two-population model to account for a series of five populations. Let us denote

these as populations A, B, C, D and E. The populations are arranged in series, and the release occurs in population A. We further assume that the mating pool in population A is made up of individuals from populations A and B, the mating pool in population B is made up of individuals from populations A, B and C, and so on. Using *Medea* as a case study, for a migration rate of μ in both directions this means that we need to make the substitutions given in eqn (18) for population A, while for population B we need to make the revised substitutions,

$$(u_{B,k}, v_{B,k}, w_{B,k}) \leftarrow (u_{B,k}, v_{B,k}, w_{B,k})(1 - 2\mu) + (u_{A,k}, v_{A,k}, w_{A,k})\mu + (u_{C,k}, v_{C,k}, w_{C,k})\mu, \quad (85)$$

Analogous substitutions apply to populations C, D and E. Considering a release of proportion x in population A at generation 0, the initial condition for the resulting difference equations is given by,

$$(u_{A,0}, w_{A,0}) = (1 - x, x), \quad (86)$$

$$(u_{B,0}, w_{B,0}) = (u_{C,0}, w_{C,0}) = (u_{D,0}, w_{D,0}) = (u_{E,0}, w_{E,0}) = 1. \quad (87)$$

Using this initial condition and the difference equations described above, we can calculate the time-series dynamics in all five populations.

Fig. 5 depicts the spread of the non-invasive gene drive systems described above through populations B through E (population A is omitted here to provide resolution for neighboring populations). Migration rates of $\mu = 0.01$ per generation and fitness costs of $s = 0.05$ are assumed (with heterozygosity $h = 0.5$), which are additive in the case of two-locus systems. For invasive systems – *Medea*, HEGs, TEs and *Wolbachia* – the transgene spreads to high frequencies in all populations. For non-invasive gene drive systems, the transgene spreads such that greater than 97% of individuals carry the transgene at the release site. Gene drive systems with high migration thresholds only display limited spread into population B, and are almost undetectable beyond this. The most persistent are translocations and two-locus engineered underdominance, which spread to 3.8% and 3.2% in population B, respectively. Moving one more population along, translocations and two-locus underdominance are diminished to equilibrium frequencies of 0.037% and 0.029%, respectively. The least invasive systems are single-allele underdominance and single-locus engineered underdominance, which persist at less than 0.01% in the directly neighboring population. The single-locus toxin-antidote systems – *Semele*, inverse *Medea* and *Merea* – persist at intermediate levels in neighboring populations and are virtually undetectable beyond Population B.

As mentioned above, killer-rescue is somewhat of a unique case. Fig. 4D depicts a single stochastic realization of the release strategy depicted in Fig. 4B for a series of five populations. Transgenic mosquitoes reach a maximum frequency of 98.0% in population A and this frequency gradually diminishes with distance, reaching 35.3% in population B, 8.7% in population C, 3.1% in population D and 1.8% in population E. The system therefore spreads the furthest and results in the largest maximum transgene frequencies in neighboring populations; however, at the same time, the system can be eliminated simply by ceasing releases provided the rescue allele is associated with a robust fitness cost. Higher fitness

costs and smaller release sizes diminish spread to neighboring populations, but also have the effect of making the system less effective at its release site.

9. Discussion

We have compared the confinement properties of a variety of gene drive systems being considered to drive disease-refractory genes into mosquito populations. Our results highlight several systems with desirable features for confinement. In simple two and five-population models, these systems require a high migration rate to become established in neighboring populations (greater than 4% per generation), and persist at low frequencies in neighboring populations for moderate migration rates (less than 5% for a migration rate of 1% per generation). Single-allele underdominance and single-locus engineered underdominance have the strongest confinement properties, reaching a prevalence of 0.01% or less in neighboring populations, and remaining confined to their release site for migration rates up to ~10% per generation for single-allele underdominance and ~18% per generation for two-locus engineered underdominance. While these systems show promise for confinement, single-allele underdominance is difficult to engineer, at least in the case where hybrids are completely unviable. Single-locus engineered underdominance requires a very high introduction frequency in order to spread at its release site (upwards of 67%); but this may not be a serious concern if the primary goal is to bring about replacement of an isolated population of modest size.

Toxin-antidote systems such as *Semele*, *Merea* and two-locus engineered underdominance also show promising confinement properties, while requiring lower introduction frequencies. *Semele*, for instance, will spread at an isolated release site for introduction frequencies greater than ~39%, and is predicted to remain confined to this population for migration rates less than ~7% per generation. For a migration rate of 1% per generation, *Semele* reaches a prevalence of less than 2% in directly neighboring populations. *Merea* displays similar confinement properties to *Semele*, but is able to spread much more quickly at its release site and can reach fixation in an isolated population despite a fitness cost. The difficulty with *Merea* is that it requires engineering a recessive antidote and, while efforts are ongoing to engineer recessive components (Hay *et al.*, unpublished), these efforts have not yet succeeded. *Semele* has also not yet been engineered, although several approaches can be taken using currently available molecular tools (Marshall *et al.*, 2011). Two-locus engineered underdominance will spread for even smaller introduction frequencies (greater than ~33%), but has a proportionately lower migration threshold of ~4% per generation. It is predicted to reach a prevalence of ~3% in neighboring populations for a migration rate of 1% per generation. Engineered underdominance requires the creation of two simultaneously-acting toxin-antidote pairs. Systems with these characteristics may be relatively easy to engineer using protein toxins and small RNAs that silence protein expression as antidotes.

Translocations have long been suggested as tools for controlling insect populations (Serebrovskii, 1940), and model predictions suggest that they can be confined to partially-isolated populations; however, for a relatively high release threshold (greater than 53% for a fitness cost of $s = 0.05$), they display a weaker strength of confinement than the above-mentioned toxin-antidote systems. Invasive gene drive systems such as HEGs and TEs show little hope of confinement. *Medea* and *Wolbachia* display moderate confinement properties in the presence of large fitness costs; however, for the case of *Medea*, it is expected to be difficult to engineer a high fitness cost in a way that is robust to mutational inactivation. Fitness costs associated with *Wolbachia* may also evolve over time, as observed in populations of *D. simulans* (Weeks *et al.*, 2007), suggesting that they cannot be relied upon for confinement. An exception is if a *Wolbachia* infection is introduced into a population that is already infected with another *Wolbachia* strain acting through a different mechanism.

In this case, the resulting bidirectional CI produces confinement properties intermediate between those of single-allele underdominance and toxin-antidote systems such as *Semele*, *Merea* and two-locus engineered underdominance.

An interesting comparison can be made between killer-rescue constructs and gene drive systems with thresholds. Killer-rescue is self-limiting in time (Gould *et al.*, 2008); but it is able to disperse through multiple populations and reach higher frequencies in neighboring populations than gene drive systems with thresholds. This does not mean that its spread is unlimited; but rather that the release site must be considered to include more communities than a release of engineered underdominance, for instance. Killer-rescue has the benefit that its presence is temporary; but it may take years for associated fitness costs to completely remove rescue transgenes from the release site and neighboring populations. In contrast, gene drive systems with thresholds can be completely removed from populations within months through a mass-release of wild-type insects, which dilute transgenes to sub-threshold frequencies.

Given the simplicity of the modeling framework, the results of this study should be interpreted as comparative rather than predictive, with an emphasis on guiding engineering efforts aimed at providing genetic tools for local population replacement. Once such tools have been developed, a detailed ecological analysis will be necessary to assess the feasibility of local replacement prior to their implementation. Such an analysis must take into account demographic details of local mosquito populations, including seasonal fluctuations in population sizes and migration rates, and variations in chromosomal form and species make-ups over time (Lanzaro *et al.*, 1998; Taylor *et al.*, 2001; Tripet *et al.*, 2005). Previous analyses have modeled asymmetries in population sizes (Marshall & Hay, 2011a) and migration rates (Altrock *et al.*, 2010), and found these to be significant. Models of invasive gene drive systems have accounted for reduced gene flow between sibling species and chromosomal forms of the *Anopheles gambiae* species complex (Taylor & Manoukis, 2003), and these considerations are even more relevant for gene drive systems with thresholds, which could potentially be confined to the chromosomal form or species of release. These points highlight the need for a deep understanding of mosquito ecology at sites where local replacement systems could eventually be released.

In summary, the existence of gene drive systems with strong confinement properties provides hope that the competing mandates of population replacement – local spread with spatial confinement – are achievable. Local replacement systems will be very useful once refractory genes have been identified, and laboratory studies and contained field trials have been completed (Benedict *et al.*, 2008). Threshold-dependent drive systems will allow disease-refractory genes to be driven to sufficiently high frequencies locally to test their hypothesized epidemiological effect (Boete & Koella, 2002; Boete & Koella, 2003), while minimizing contamination of immediately neighboring populations. This is an important step towards gaining acceptance for the use of invasive gene drive systems, with implications for both disease control and biosafety on a global scale.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Catherine Ward for advice on modeling, Charles Taylor for helpful discussions on mosquito dispersal, and the editor and two anonymous reviewers whose constructive comments have greatly improved the manuscript. John M. Marshall was supported by grant number DP1 OD003878 to Bruce A. Hay from the National Institutes of Health, and by a grant from the Medical Research Council, UK.

References

- Alphey L, Beard CB, Billingsley P, Coetzee M, Crisanti A, et al. Malaria control with genetically manipulated insect vectors. *Science*. 2002; 298:119–121. [PubMed: 12364786]
- Altrock PM, Traulsen A, Reeves RG, Reed FA. Using underdominance to bistably transform local populations. *J Theor Biol*. 2010; 267:62–75. [PubMed: 20691703]
- Barton NH. The dynamics of hybrid zones. *Heredity*. 1979; 43:341–359.
- Beech CJ, Nagaraju J, Vasani SS, Rose RI, Othman RY, et al. Risk analysis of a hypothetical open field release of a self-limiting transgenic *Aedes aegypti* mosquito strain to combat dengue. *AsPac J Mol Biol Biotechnol*. 2009; 17:99–111.
- Beeman RW, Friesen KS. Properties and natural occurrence of maternal-effect selfish genes ('Medea' factors) in the red flour beetle, *Tribolium castaneum*. *Heredity*. 1999; 82:529–534. [PubMed: 10383673]
- Beeman RW, Friesen KS, Denell RE. Maternal-effect selfish genes in flour beetles. *Science*. 1992; 256:89–92. [PubMed: 1566060]
- Benedict M, D'Abbs P, Dobson S, Gottlieb M, Harrington L, et al. Guidance for contained field trials of vector mosquitoes engineered to contain a gene drive system: recommendations of a scientific working group. *Vector Borne Zoonotic Dis*. 2008; 8:127–166. [PubMed: 18452399]
- Boete C, Koella JC. A theoretical approach to predicting the success of genetic manipulation of malaria mosquitoes in malaria control. *Malaria J*. 2002; 1:3.
- Boete C, Koella JC. Evolutionary ideas about genetically manipulated mosquitoes and malaria control. *Trends Parasitol*. 2003; 19:32–38. [PubMed: 12488224]
- Braig, HR.; Yan, G. The spread of genetic constructs in natural insect populations. In: Letourneau, DK.; Burrows, BE., editors. *Genetically Engineered Organisms: Assessing Environmental and Human Health Effects*. CRC Press; New York: 2001. p. 251–314.
- Burt A. Site-specific genes as tools for the control and genetic engineering of natural populations. *Proc Biol Sci*. 2003; 270:921–928. [PubMed: 12803906]
- Bush GL, Case SM, Wilson AC, Patton JL. Rapid speciation and chromosomal evolution in mammals. *Proc Natl Acad Sci USA*. 1977; 74:3942–3946. [PubMed: 269445]
- Catteruccia F, Godfray HCJ, Crisanti A. Impact of genetic manipulation on the fitness of *Anopheles stephensi* mosquitoes. *Science*. 2003; 299:1225–1227. [PubMed: 12595691]
- Charlesworth B, Charlesworth D. The population dynamics of transposable elements. *Genet Res*. 1983; 42:1–27.
- Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*. 1994; 371:215–220. [PubMed: 8078581]
- Chen CH, Huang H, Ward CM, Su JT, Schaeffer LV, et al. A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*. *Science*. 2007; 316:597–600. [PubMed: 17395794]
- Curtis CF. Possible use of translocations to fix desirable genes in insect pest populations. *Nature*. 1968; 218:368–369. [PubMed: 5649682]
- Curtis CF, Robinson AS. Computer simulation of the use of double translocations for pest control. *Genetics*. 1971; 69:97–113. [PubMed: 5167305]
- Davis S, Bax N, Grewe P. Engineered underdominance allows efficient and economical introgression of traits into pest populations. *J Theor Biol*. 2001; 212:83–98. [PubMed: 11527447]
- Deredec A, Burt A, Godfray HC. The population genetics of using homing endonuclease genes in vector and pest management. *Genetics*. 2008; 179:2013–2026. [PubMed: 18660532]
- Egli D, Hafen E, Schaffner W. An efficient method to generate chromosomal rearrangements by targeted DNA double-strand breaks in *Drosophila melanogaster*. *Genome Research*. 2004; 14:1382–1393. [PubMed: 15197166]
- Elaydi, SN. *An Introduction to Difference Equations*. Springer; New York: 1995.
- Engelstadter J, Telschow A. Cytoplasmic incompatibility and host population structure. *Heredity*. 2009; 103:196–207. [PubMed: 19436325]

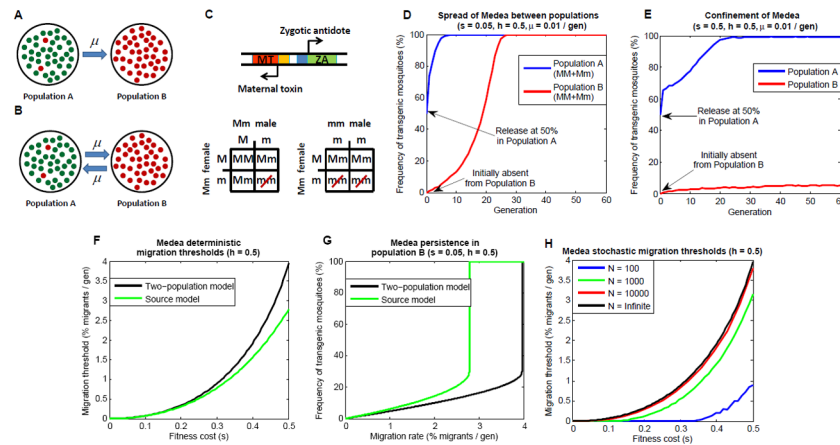
- Flor M, Hammerstein P, Telschow A. Dynamics and stability of *Wolbachia*-induced unidirectional cytoplasmic incompatibility in parapatric host populations. *J Evol Biol.* 2007; 20:696–706. [PubMed: 17305835]
- Foster GG, Whitten MJ, Prout T, Gill R. Chromosome rearrangements for the control of insect pests. *Science.* 1972; 176:875–880. [PubMed: 4402399]
- Golic KG, Golic MM. Engineering the *Drosophila* genome: Chromosome rearrangements by design. *Genetics.* 1996; 144:1693–1711. [PubMed: 8978056]
- Gould F, Schliekelman P. Population genetics of autocidal control and strain replacement. *Annu Rev Entomol.* 2004; 49:193–217. [PubMed: 14651462]
- Gould F, Huang Y, Legros M, Lloyd AL. A killer-rescue system for self-limiting gene drive of anti-pathogen constructs. *Proc Roy Soc B.* 2008; 275:2823–2829.
- Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, et al. Reducing *Plasmodium falciparum* malaria transmission in Africa: A model-based evaluation of intervention strategies. *PLoS Med.* 2010; 7:e1000324.
- Guimond N, Bideshi DK, Pinkerton AD, Atkinson PW, O'Brochta DA. Patterns of *Hermes* transposition in *Drosophila melanogaster*. *Molec Gen Genet.* 2003; 268:779–790.
- Hartl, DL.; Clark, AG. Principles of Population Genetics. Sinauer Associates, Inc; Sunderland, MA: 1997.
- Hoffmann AA, Turelli M, Harshman LG. Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. *Genetics.* 1990; 126:933–948. [PubMed: 2076821]
- Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaeate I, Johnson PH, et al. Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature.* 2011; 476:454–457. [PubMed: 21866160]
- Jansen VAA, Turelli M, Godfray HCJ. Stochastic spread of *Wolbachia*. *Proc Roy Soc London Ser B.* 2008; 275:2769–2776.
- Karlin S, McGregor J. Application of method of small parameters to multi-niche population genetic model. *Theor Popul Biol.* 1972; 3:186–209. [PubMed: 4667083]
- Keeling MJ, Jiggins FM, Read JM. The invasion and coexistence of competing *Wolbachia* strains. *Heredity.* 2003; 91:382–388. [PubMed: 14512953]
- Kidwell MG. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc Natl Acad Sci USA.* 1983; 80:1655–1659. [PubMed: 6300863]
- Knols BG, Bossin HC, Mukabana WR, Robinson AS. Transgenic mosquitoes and the fight against malaria: managing technology push in a turbulent GMO world. *Am J Trop Med Hyg.* 2007; 77:232–242. [PubMed: 18165498]
- Lande R. Effective deme size during long term evolution estimated from rates of chromosomal rearrangements. *Evolution.* 1979; 33:234–251.
- Lande R. The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity.* 1985; 54:323–332. [PubMed: 4019220]
- Lanzaro GC, Toure YT, Carnahan J, Zheng L, Dolo G, et al. Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proc Natl Acad Sci USA.* 1998; 95:14260–14265. [PubMed: 9826688]
- Le Rouzic A, Capi P. The first steps of a transposable element invasion: Parasitic strategy vs. drift. *Genetics.* 2005; 169:1033–1043. [PubMed: 15731520]
- Lorenzen MD, Gnirke A, Margolis J, Garnes J, Campbell M, et al. The maternal-effect, selfish genetic element *Medea* is associated with a composite Tc1 transposon. *Proc Natl Acad Sci USA.* 2008; 105:10085–10089. [PubMed: 18621706]
- Lorimer N, Hallinan E, Rai KS. Translocation homozygotes in the yellow fever mosquito, *Aedes aegypti*. *J Hered.* 1972; 63:158–166. [PubMed: 4672742]
- Mackay TF, Lyman RE, Jackson MS. Effects of *P*-element insertions on quantitative traits in *Drosophila melanogaster*. *Genetics.* 1992; 130:315–332. [PubMed: 1311697]
- Magori K, Gould F. Genetically engineered underdominance for manipulation of pest populations: a deterministic model. *Genetics.* 2006; 172:2613–2620. [PubMed: 16415364]

- Marshall JM. The impact of dissociation on transposon-mediated disease control strategies. *Genetics*. 2008; 178:1673–1682. [PubMed: 18245837]
- Marshall JM. The effect of gene drive on containment of transgenic mosquitoes. *J Theor Biol*. 2009; 258:250–265. [PubMed: 19490857]
- Marshall JM. The Cartagena Protocol and genetically modified mosquitoes. *Nat Biotech*. 2010; 28:896–897.
- Marshall JM. The toxin and antidote puzzle: New ways to control insect pest populations through manipulating inheritance. *Bioeng Bugs*. 2011; 2:1–6. [PubMed: 21636981]
- Marshall JM, Hay BA. Inverse *Medea* as a novel gene drive system for local population replacement: A theoretical analysis. *J Hered*. 2011a; 102:336–341. [PubMed: 21493596]
- Marshall JM, Hay BA. General principles of single-construct chromosomal gene drive. 2011b (under review).
- Marshall JM, Taylor CE. Malaria control with transgenic mosquitoes. *PLoS Medicine*. 2009; 6:e20. [PubMed: 19209953]
- Marshall JM, Toure MB, Traore MM, Famenini S, Taylor CE. Perspectives of people in Mali toward genetically-modified mosquitoes for malaria control. *Malaria J*. 2010; 9:128.
- Marshall JM, Pittman GW, Buchman AB, Hay BA. *Semele*: A killer-male rescue-female system for suppression and replacement of insect disease vector populations. *Genetics*. 2011; 187:535–551. [PubMed: 21078687]
- Maside X, Assimakopoulos S, Charlesworth B. Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. *Genet Res*. 2000; 75:275–284. [PubMed: 10893864]
- Nuzhdin SV, Pasyukova EG, Mackay TFC. Accumulation of transposable elements in laboratory lines of *Drosophila melanogaster*. *Genetica*. 1997; 100:167–175. [PubMed: 9440270]
- Pinto J, Donnelly MJ, Sousa CA, Malta-Vacas J, Gil V, et al. An island within an island: genetic differentiation of *Anopheles gambiae* in Sao Tome, West Africa, and its relevance to malaria vector control. *Heredity*. 2003; 91:407–414. [PubMed: 14512957]
- Rasgon JL, Gould F. Transposable element insertion location bias and the dynamics of gene drive in mosquito populations. *Insect Mol Biol*. 2005; 14:493–500. [PubMed: 16164605]
- Reuter M, Lehmann L, Guillaume F. The spread of incompatibility-inducing parasites in sub-divided host populations. *BMC Evol Biol*. 2008; 8:134. [PubMed: 18460188]
- Ribeiro JMC, Kidwell MG. Transposable elements as population drive mechanisms: Specification of critical parameter values. *J Med Entomol*. 1994; 31:10–16. [PubMed: 8158612]
- Robinson AS. Progress in the use of chromosomal translocations for the control of insect pests. *Biol Rev*. 1976; 51:1–24. [PubMed: 3229]
- Schofield P. Spatially explicit models of Turelli-Hoffmann *Wolbachia* invasive wave fronts. *J Theor Biol*. 2002; 215:121–131. [PubMed: 12051989]
- Seleme M, Busseau I, Malinsky S, Bucheton A, Teninges D. High-frequency retrotransposition of a marked *I* factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. *Genetics*. 1999; 151:761–771. [PubMed: 9927467]
- Serebrovskii AS. On the possibility of a new method for the control of insect pests. *Zool Zh*. 1940; 19:618–690.
- Service, MW. *Mosquito Ecology: Field Sampling Methods*. Chapman and Hall; London: 1993.
- Sinkins SP, Gould F. Gene drive systems for insect disease vectors. *Nat Rev Genet*. 2006; 7:427–435. [PubMed: 16682981]
- Stouthamer R, Breeuwer JA, Hurst GD. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annu Rev Microbiol*. 1999; 53:71–102. [PubMed: 10547686]
- Struchiner CJ, Massad E, Tu Z, Ribeiro JMC. The tempo and mode of evolution of transposable elements as revealed by molecular phylogenies reconstructed from mosquito genomes. *Evolution*. 2009; 63:3136–3146. [PubMed: 19656180]
- Subramanian RA, Arensburger AP, Atkinson PW, O'Brochta DA. Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector *Anopheles gambiae* s. s. *Genetics*. 2007; 176:2477–2487. [PubMed: 17603116]

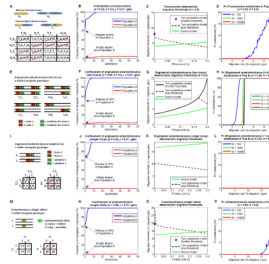
- Taylor, CE.; Manoukis, NC. Effective population size in relation to genetic modification of *Anopheles gambiae* sensu stricto. In: Takken, W.; Scott, TW., editors. Ecological Aspects for Application of Genetically Modified Mosquitoes. Wageningen; The Netherlands: 2003. p. 133-146.
- Taylor CE, Toure YT, Carnahan J, Norris DE, Dolo G, et al. Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics*. 2001; 157:743–750. [PubMed: 11156993]
- Telschow A, Yamamura N, Werren JH. Bidirectional cytoplasmic incompatibility and the stable coexistence of two *Wolbachia* strains in parapatric host populations. *J Theor Biol*. 2005; 235:265–274. [PubMed: 15862594]
- Tower J, Karpen GH, Craig N, Spradling AC. Preferential transposition of *Drosophila* *P* elements to nearby chromosomal sites. *Genetics*. 1993; 133:347–359. [PubMed: 8382177]
- Townsend JP, Hartl DL. The kinetics of transposable element autoregulation. *Genetica*. 2000; 108:229–237. [PubMed: 11294609]
- Tripet F, Dolo G, Lanzaro G. Multilevel analyses of genetic differentiation in *Anopheles gambiae* s.s reveal patterns of gene flow important for malaria-fighting mosquito projects. *Genetics*. 2005; 169:313–314. [PubMed: 15677750]
- Turelli M. Cytoplasmic incompatibility in populations with overlapping generations. *Evolution*. 2010; 64:232–241. [PubMed: 19686264]
- Turelli M, Hoffmann AA. Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature*. 1991; 353:440–442. [PubMed: 1896086]
- Turelli M, Hoffmann AA. Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. *Insect Mol Biol*. 1999; 8:243–255. [PubMed: 10380108]
- Vasilyeva LA, Bubenshchikova EV, Ratner VA. Heavy heat shock induced retrotransposon transposition in *Drosophila*. *Genet Res*. 1999; 74:111–119. [PubMed: 10584555]
- Wade MJ, Beeman RW. The population dynamics of maternal-effect selfish genes. *Genetics*. 1994; 138:1309–1314. [PubMed: 7896109]
- Wade MJ, Stevens L. The effect of population subdivision on the rate of spread of parasite-mediated cytoplasmic incompatibility. *J Theor Biol*. 1994; 167:81–87. [PubMed: 7909859]
- Walker T, Johnson PH, Moreira LA, Iturbe-Ormaeche I, Frentiu FD, et al. The wMel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature*. 2011; 476:450–453. [PubMed: 21866159]
- Ward CM, Su JT, Huang Y, Lloyd AL, Gould F, et al. *Medea* selfish genetic elements as tools for altering traits of wild populations: A theoretical analysis. *Evolution*. 2010; 64:1149–1162. [PubMed: 21062278]
- Weeks AR, Turelli M, Harcombe WR, Reynolds KT, Hoffmann AA. From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila*. *PLoS Biology*. 2007; 5:e114. [PubMed: 17439303]
- Windbichler N, Papathanos PA, Crisanti A. Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in *Anopheles gambiae*. *PLoS Genet*. 2008; 4:e1000291. [PubMed: 19057670]
- Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, et al. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature*. 2011; 473:212–215. [PubMed: 21508956]
- World Health Organization. World Malaria Report 2010. WHO; Geneva, Switzerland: 2010.
- Wright S. On the probability of fixation of reciprocal translocations. *Am Nat*. 1941; 75:513–522.
- Yeap HL, Mee P, Walker T, Weeks AR, O'Neill SL, et al. Dynamics of the 'popcorn' *Wolbachia* infection in outbred *Aedes aegypti* informs prospects for mosquito vector control. *Genetics*. 2010; 185:1534.
- Zhang P, Spradling AC. Efficient and dispersed local *P* element transposition from *Drosophila* females. *Genetics*. 1993; 133:361–373. [PubMed: 8382178]

Highlights

- Gene drive systems are planned to drive anti-disease genes into vector populations.
- Confinement of transgenes to release sites is desirable during field trials.
- Underdominant systems spread locally and show strong confinement potential.
- Toxin-antidote systems are confineable and require lower introduction frequencies.
- Killer-rescue is self-limiting in time but disperses into neighboring populations.

**Fig. 1.**

Dynamics of *Medea* under the source and two-population models. (A): In the source model, we assume that transgenic mosquitoes have already reached equilibrium in population A. Population A remains at equilibrium and donates a fraction, μ , of its population to population B at each generation. (B): In the two-population model, each population exchanges a fraction, μ , of its population with the other at each generation. Transgenic mosquitoes are introduced into population A. Population B is initially wild-type. (C): *Medea* is comprised of two genes – a maternal toxin and zygotic antidote. The expression of these genes renders wild-type offspring of heterozygous females unviable. (D): Single stochastic realization of a *Medea* element released at 50% (in the form of *Medea* homozygotes) in population A. Population A exchanges migrants with population B at a rate of 1% per generation and the size of both populations is 10,000. The *Medea* construct is associated with a fitness cost of $s = 0.05$ ($h = 0.5$). *Medea*-bearing genotypes fix in both populations within 27 generations. (E): The element is associated with a fitness cost of $s = 0.5$ ($h = 0.5$). *Medea* reaches near-fixation in population A within 20 generations, but only spreads to ~4.7% (in the form of *Medea* homozygotes and heterozygotes) in population B. (F): *Medea* displays threshold behavior with respect to migration rate. For a given fitness cost, there is a migration rate above which *Medea* fixes and below which it merely persists at a low level in population B. (G): *Medea* persistence in population B depends on migration rate between the two populations. (H): In stochastic systems, there is a chance that *Medea* will become established in population B even when the migration rate is below the deterministic threshold. We define the stochastic threshold as the migration rate above which the probability that the element becomes established in the neighboring population (i.e. spreads to a transgenic frequency above 90%) within 1,000 generations is greater than 0.1%.

**Fig. 2.**

Dynamics of underdominant systems under the source and two-population models. (A): Reciprocal translocations form when terminal segments of two nonhomologous chromosomes are mutually exchanged. These display underdominant dynamics since, when a translocation heterozygote undergoes meiosis, half its gametes are unbalanced and many combinations of gametes produce unviable offspring. (B): Single stochastic realization of a translocation released at 60% in population A. The translocation is associated with a fitness cost of $s=0.05$ ($h=0.5$). Population A exchanges migrants with population B at a rate of 1% per generation and the size of both populations is 10,000. The translocation reaches near-fixation in population A within 15 generations but only spreads to ~4.2% in population B. (C): For the source model, the translocation displays threshold behavior with respect to migration rate. For the two-population model, there is a migration threshold below which the translocation fixes in population A and persists at a low level in population B, and above which it is lost in both populations. (D): For small population sizes (~100 individuals) and migration rates above 4% per generation, there is a chance that the translocation will become established in population B under the two-population model. (E): Engineered underdominance consists of two constructs, each of which possesses a toxin gene and an antidote gene that suppresses the expression or activity of the toxin gene carried by the other construct. In the two-locus system, the constructs are inserted at distant loci on the same chromosome or at different loci on nonhomologous chromosomes. (F): Single stochastic realization of constructs released at 50% in population A. Population A exchanges migrants with population B at a rate of 1% per generation and the size of both populations is 10,000. The two constructs are associated with additive fitness costs of $s=0.05$ ($h=0.5$). The constructs reach near-fixation in population A within 10 generations, but only spread to a frequency of ~3.2% in population B (in the form of individuals having at least one of each transgenic construct). (G): Two-locus underdominance displays threshold behavior with respect to migration rate. (H): As population size decreases, there is a chance that two-locus underdominance can become established in neighboring populations for smaller migration rates. (I): For single-locus engineered underdominance, the toxin-antidote constructs are inserted at the same location on homologous chromosomes. (J): Single stochastic realization of constructs released at 70% in population A. Population A exchanges migrants with population B at a rate of 1% per generation and the size of both populations is 10,000. The constructs are associated with a fitness cost of $s=0.05$. The constructs reach near-fixation in population A within eight generations but only spread to ~0.005% in population B. (K): For the source model, single-locus engineered underdominance displays threshold behavior with respect to migration rate. For the two-population model, there is a migration threshold below which the constructs fix in population A and persist at a low level in population B, and above which they are lost in both populations. (L): Even for stochastic populations with 100 individuals, single-locus engineered underdominance is not expected to become established in population B. (M): An underdominant allele for which hybrids are completely unviable. (N): Single stochastic realization of an allele released at 60% in population A. Population A exchanges migrants with population B at a rate of 1% per generation and the size of both populations is 10,000. For a fitness cost of $s=0.05$, the construct reaches near-fixation in

population A within five generations but only spreads to ~0.01% in population B. (O): For the source model, extreme underdominance displays threshold behavior with respect to migration rate. For the two-population model, there is a migration threshold below which the construct fixes in population A and persists at a low level in population B and above which it is lost in both populations. (P): For small population sizes (~100 individuals) and migration rates above 15% per generation, there is a chance that the underdominant allele will become established in both populations under the two-population model.

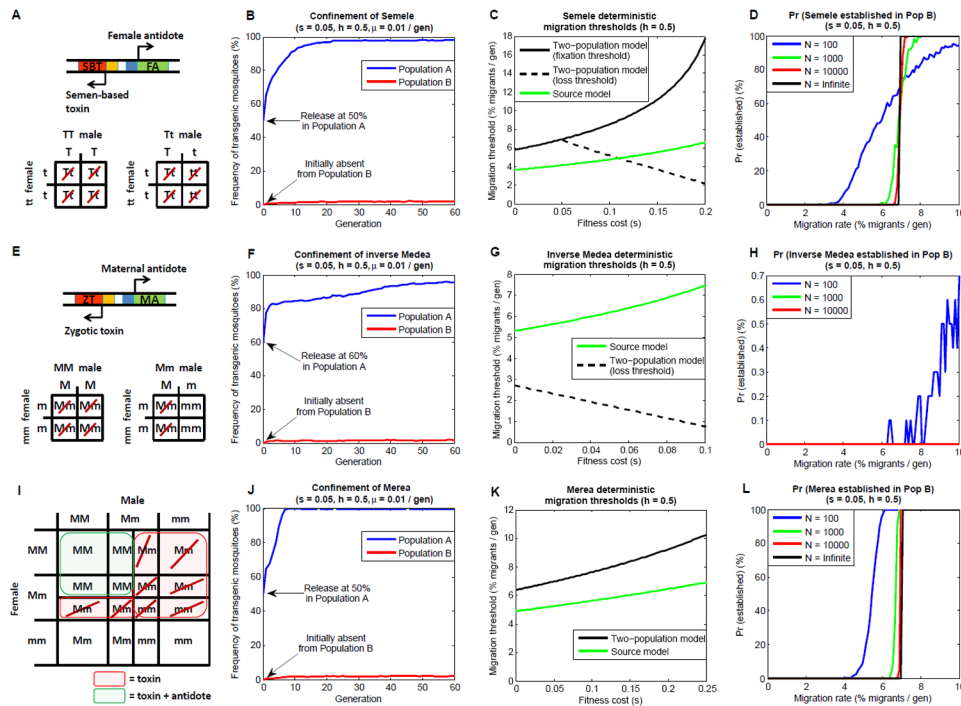
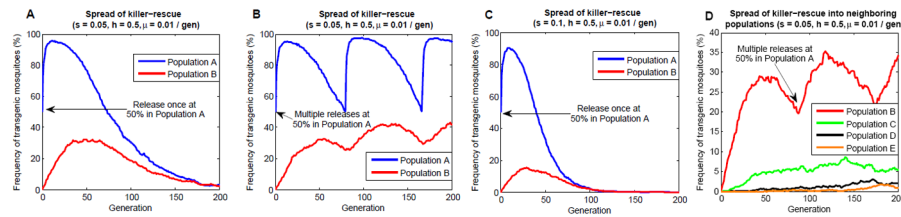


Fig. 3.

Dynamics of other non-invasive gene drive systems under the source and two-population models. (A): *Semele* consists of a toxin expressed in the semen of transgenic males and an antidote expressed in females. This causes transgenic females to be favored at high population frequencies. (B): Single stochastic realization of an element released at 50% in population A. Population A exchanges migrants with population B at a rate of 1% per generation and the size of both populations is 10,000. The element is associated with a fitness cost of $s = 0.05$ ($h = 0.5$). *Semele* reaches near-fixation in population A within 20 generations but only spreads to ~1.8% in population B. (C): *Semele* displays threshold behavior with respect to migration rate. For the two-population model, there is a migration threshold above which the element reaches near-fixation in both populations and, in some cases, a range of smaller migration rates for which the element is lost from both populations. (D): For smaller population sizes, there is a chance that *Semele* will become established in neighboring populations for smaller migration rates. (E): *Inverse Medea* consists of a zygotic toxin and maternal antidote which render heterozygous offspring of wild-type mothers unviable. (F): Single stochastic realization of an element released at 60% in population A. Population A exchanges migrants with population B at a rate of 1% per generation and the size of both populations is 10,000. The element is associated with a fitness cost of $s = 0.05$ ($h = 0.5$). *Inverse Medea* quickly spreads to 80% in population A and then slowly spreads to near-fixation within 60 generations, but only spreads to ~1.6% in population B. (G): For the source model, *Inverse Medea* displays threshold behavior with respect to migration rate. For the two-population model, there is a migration threshold below which the element fixes in population A and persists at a low level in population B, and above which it is lost in both populations. (H): For small population sizes (~100 individuals) and migration rates above 6% per generation, there is a chance that *Inverse Medea* will become established in both populations under the two-population model. (I): *Merea* consists of a maternal toxin and recessive zygotic antidote. Heterozygous offspring are not rescued by the antidote and hence the element must exceed a threshold frequency in order to distort the offspring ratio in its favor. (J): Single stochastic realization of an element released at 50% in population A. Population A exchanges migrants with population B at a rate of 1%

per generation and the size of both populations is 10,000. The element is associated with a fitness cost of $s=0.05$ ($h=0.5$). *Merea* fixes in population A within eight generations but only spreads to ~2.0% in population B. (K): *Merea* displays threshold behavior with respect to migration rate. (L): For smaller population sizes, there is a chance that *Merea* will become established in neighboring populations for smaller migration rates.

**Fig. 4.**

Stochastic time-series dynamics of killer-rescue constructs released at a frequency of 50% in population A. Both alleles (K and R) are associated with additive fitness costs of $s = 0.05$ ($h = 0.5$). Population A exchanges migrants with population B at each generation. The size of both populations is 10,000 and the migration rate between the two populations is 1% per generation. (A): For a single release, transgenic mosquitoes reach a peak frequency of 95% within 11 generations and decline in frequency over the next 200 generations. In population B, transgenic mosquitoes reach a peak frequency of 29% within 43 generations and decline in frequency over the next 150 generations. (B): For a release every time transgenic frequency falls below 50%, transgenic mosquitoes reach a peak frequency of 98% after the second release and oscillate between this and 50%. In population B, transgenic mosquitoes reach a peak frequency of 43% after the third release and oscillate between this and 31%. (C): If the additive fitness costs are increased to $s = 0.1$ ($h = 0.5$), killer-rescue contaminates neighboring populations less; but also spreads to lower frequencies at its release site. For a release every time transgenic frequency falls below 50%, transgenic mosquitoes reach a peak frequency of 94% after the second release and oscillate between this and 50%. In population B, transgenic mosquitoes reach a peak frequency of 23% after the third release and oscillate between this and 16%. (D): Spread of killer-rescue through a series of five neighboring populations for the release strategy depicted in panel B. Killer-rescue peaks at 35.3% in population B and can be found at frequencies as high as 1.8% in population E.

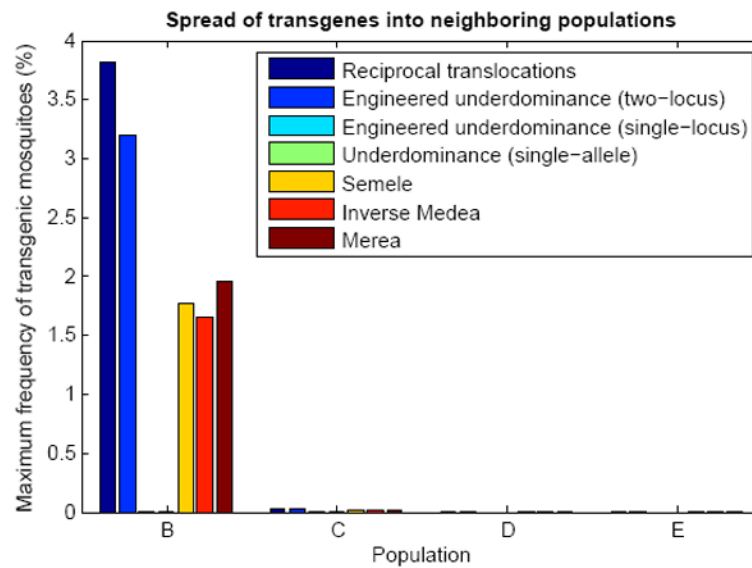


Fig. 5.

Spread of transgenes linked to gene drive systems through a series of five neighboring populations. Each transgenic allele has a fitness cost of $s = 0.05$. Equilibrium transgenic frequencies are shown for a super-threshold release in population A. In all cases, the transgene spreads to a frequency above 97% in population A (not shown), plateaus at a frequency below 5% in population B, and is almost undetectable in populations C–E.

TABLE 1

Parameter values, definitions and source references.

Parameter:	Definition:	Default value:	References:
μ	Migration rate	0.01/gen	Taylor <i>et al.</i> (2001); Tripet <i>et al.</i> (2005)
N	Population size	10,000	
<u>Medea, HEG, Semele, inverse Medea, Merea, extreme underdominance, engineered underdominance (single-locus, two-locus), translocations, killer-rescue:</u>			
s	Homozygous fitness cost (per locus)	0.05	Chen <i>et al.</i> (2007); Catteruccia <i>et al.</i> (2003)
h	Heterozygosity of fitness cost	0.5	
<u>TE:</u>			
u_1	Replicative transposition rate	0.1/gen	Seleme <i>et al.</i> (1999); Vasilyeva <i>et al.</i> (1999)
λ	Transpositional regulation parameter	1	Subramanian <i>et al.</i> (2007)
v	Element deletion rate	4×10^{-6} /gen	Nuzhdin <i>et al.</i> (1997); Maside <i>et al.</i> (2000)
s	Fitness cost per TE	0.01	Mackay <i>et al.</i> (1992)
<u>HEG:</u>			
e	Homing rate	0.1/gen	Deredec <i>et al.</i> (2008)
<u>Wolbachia, bidirectional CI:</u>			
u	Degree of imperfection in maternal transmission	0.97	Turelli & Hoffmann (1999)
e	Efficiency of CI-induced sterility	0.7	
s	Fitness cost of <i>Wolbachia</i>	0.05	

TABLE 2

Comparison of gene drive systems in terms of confinement properties.*

Gene drive system:	Migration threshold (/gen) (source model):	Migration threshold (/gen) (2-pop model):	Stochastic migration threshold (/gen) (2-pop model, N=100/1000):	Transgenic frequency in pop B (2-pop model, 1% migrants/gen):	Mean time to establishment in Pop B (gen) (2-pop model, N=1000, 1% migrants/gen & at migration threshold):**
<i>Medea</i>	0.017%	0.018%	0%/0%	100%	21.8 & 397.1
HEG	0%	0%	0%/0%	100%	37.7 & –
TE	0%	0%	0%/0%	100%	37.1 & –
<i>Wolbachia</i>	0.24%	0.25%	0%/0.055%	100%	35.03 & 234.4
Bidirectional CI	8.8%	13.3%	7.0%/11.2%	0.5%	– & 45.9
Translocations	4.3%	–	2.7%/6.0%	4.2%	– & –
Engineered underdominance (two-locus)	2.8%	4.3%	1.5%/3.7%	3.2%	– & 119.2
Engineered underdominance (single-locus)	17.0%	–	–/–	0.005%	– & –
Extreme underdominance	12.8%	–	14.7%/19.4%	0.01%	– & –
<i>Semele</i>	4.2%	6.9%	2.7%/6.0%	1.8%	– & 46.4
Inverse <i>Medea</i>	6.2%	–	6.4%/–	1.6%	– & –
<i>Merea</i>	5.3%	7.0%	4.0%/6.5%	2.0%	– & 109.0
Killer-rescue	0%	0%	0%/0%	–	– & –

* Default parameter values are taken from Table 1.

** Initial prevalence in Population A is 100%.